Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2016-399-RC2, 2017

Atmospheric Measurement Techniques
Discussions

Interactive comment on "Smoothing data series by means of cubic splines: quality of approximation and introduction of an iterative spline approach" *by* Sabine Wüst et al.

Anonymous Referee #1

Received and published: 21 March 2017

The paper by Wüst et al. deals with the use of cubic splines for smoothing of data series that contain a superposition of (different) gravity waves and a background state. The authors describe the dependence of the quality of the spline fit on the number and location of sampling points in relation to the wavelength and phase of the wave, respectively. In the atmosphere the wavelengths of the observed waves are typically not well known before detailed analysis and not constant across the data set. A distinct phase relation may in this case result in large approximation errors of the spline fit. The authors therefore propose an "iterative" approach with variable phase relation. The new method is applied to temperature data from TIMED/SABER and used to derive gravity wave activities. The manuscript is well written and the figures are instructive. On

the other hand there is some confusion in the structure and content that needs to be resolved before publication. I describe my concerns in detail below.

Major comments

As the authors describe, splines are mainly used for smoothing of a data series. This can be done to extract a "background state" like a temperature profile undisturbed by gravity waves, to produce a smooth data set including only a subset of waves, or to retrieve the residuals that are then treaded as a wave disturbance and examined further. The purpose of spline fitting largely determines the spline parameters and the quantities that should be evaluated. From my point of view the actual manuscript mixes up between a spline for optimal description of a background (and extraction of waves as residuals) and for optimal wave description (minimizing residuals). For example:

- In line 11 (page 1) the authors mention the approximation of undisturbed conditions, but in line 14 they discuss effects of increasing number of sampling points, even if this obviously contradicts the purpose of getting a background state.

- In line 4/5 (page 2) the authors indeed state the goal of getting a "sufficiently low number of sampling points" to describe the background.

- In Section 3 a much shorter sampling point interval compared to the Introduction is used. Accordingly a wave with small vertical wavelength is examined. Even if residuals between the spline fit and the original are calculated, they are minimized ("approximation error") in order to get an optimal reproduction of the wave.

- In the case with CIRA and superposed waves (Section 3 and Figure 7) again a very small sampling point distance is used in the example. If the spline shall be used to extract an "undisturbed" profile, I suggest to calculate the squared difference between the fitted and the original CIRA profile, not between the fitted and the disturbed profile.

- In section 4 (p. 7, l. 2) again the "ability of a spline to approximate oscillations" is described, not the smoothing of the oscillations to retrieve the background.

- In Fig. 8, again a larger sampling point distance is applied and the residuals are examined.

- In the Summary (p. 9, l. 8) the "ability of a spline to approximate oscillations" is mentioned, not the ability to reproduce a background state (and wave-induced residuals).

Please clarify.

Minor comments:

- Introduction: I am surprised that the 10 km oscillations appear so steadily in the averaged data. The later analysis suggests that there is a gravity wave with fixed wavelength and phase that comes out in all altitudes and in all years. This is quite unexpected. Effects of changing temperature gradients (p. 8, l. 20) should appear locally in the profiles.

- P. 3, l. 16-18: This sentence structure is rather complicate. I suggest rephrasing and, e.g., separating in two sentences.

- Section 2 (and others): I find the term "iterative" misleading. Typically it is used for some processes where the adaptive change of parameters produces some converging results. What is done here is that the algorithm averages across all phase differences that are possible between sampling points and the shortest resolvable wave.

- Figure 7 b: It would be interesting to see the error also for the non-interative spline fit.

- P. 7, l. 5: If the wavelength of the sine is so close to the shortest resolvable wavelength, this dependence is not surprising. The problem should be much smaller, if less sampling points are applied and the spline is used to extract the residuals.

- P. 7, l. 25: Is it intended to compare the non-iterative and iterative method (Fig. 3 and 6a) or constant or CIRA background (Fig. 7) as suggested in l. 22. For Fig. 7 it should be noted that the scaling of error and sampling point distance differs from the other plots.

- P. 8, l. 1: I do not see that wave activity is less variable compared to higher altitudes. I only see that amplitudes show a much smaller increase with altitude compare to below and above.

- Section 4: In general it would be helpful to see the results of standard and iterative spline fitting also for a single SABER profile, containing a superposition of background and different waves. This may help to understand the mean profiles of non-squared and squared residuals.

- P. 8, l. 17-18: i) It is still surprising that the annual mean residuals ($\sim$500 profiles) show such a pronounced oscillation. Please comment on this. ii) The amplitudes shown in Fig. 8 c) and d) are non-squared averages and the individual residuals should me much larger. On the other hand the iterative approach should have largest effect if the wavelength is close to double the sampling point distance. This should mainly appear in cases of wave approximation and less in cases of background approximation.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-399, 2017.

Answer to

Interactive comment on “Smoothing data series by means of cubic splines: quality of approximation and introduction of an iterative spline approach” by Sabine Wüst et al.

First of all, we would like to thank you for your valuable comments.
In order to facilitate further discussion, we point out numbers of pages of paragraphs in our answer refer to the version with accepted changes as long as not stated otherwise.
The attached manuscript with marked changes includes also the changes due to the comments of the other reviewer.

Major comments:

The definition of background or undisturbed background in our sense depends on the addressed research question. It could contain parts of the wave spectra when one focuses one smaller-scale waves for example; in this case one uses only the residuals. Further spectral analysis requires large-scale de-trending in advance (large-scale in relation to the signatures one is interested in), and splines are a common used method for this because they approximate linear as well as wave-like structures. On the other hand, the residuals could contain small-scale non-wave disturbances which disturb the analysis of a specific scientific question and one is only interested in the spline approximation.

We re-formulated the manuscript (see below) to make clear that we are mainly interested in the residuals in common. Nevertheless, one can also learn more about the spline approximation even if the residuals are in the focus.
Additional and in accordance with the comment of the other reviewer, we also avoid the term “approximation error” now in the manuscript which might have contributed to a lack of clearness and use “sum of squared residuals”.

1. Former page 1, line 11 & 14: We changed the first paragraph of the introduction to make clear that in general the background state or the residuals—depending on the research question—can be of interest.

2. Former page 2, line 4/5: sentence deleted

3. Section 3: The values for the case studies (without background and with CIRA background) are more or less arbitrarily chosen, however we motivated their choice in the manuscript.
   The effects of the traditional spline approximation are demonstrated in this section for short (3 km) and long (up to 13 km) vertical wavelength (case study without background and with CIRA background) as well as for different sampling point distances which are also comparable with the one chosen in figure 1.

4. Since we do not focus on the optimization of the fitting per se, your comment concerning the CIRA example is obsolete. We just would like to demonstrate that the method is able to filter for a specific part of the wave spectrum. The steps of decrease in the residuals agree very well with the wavelengths.

5. We included “and therefore to filter for a specific part of the wave spectrum” in order to avoid confusion.

6. Since this figure corresponds to figure 1, we choose a comparable sampling point distance.

7. We hope that we have clarified all confusing points mentioned above and leave the formulation in the summary as it was since this meets the purpose of the manuscript.

Minor comments:

1. We interpret this 10 km oscillation as an artificial effect of the analysis (p.7 l.28ff and page 9, first paragraph). A persistent gravity wave with fixed wavelength and phase seems also to us very unrealistic.

2. Done.

3. We use the term “iterative“ in its original sense (latin: iterare = to repeat), however, we understand your comment that iterative is frequently used today not only for something which is repeated but for something which converges due to adapting certain parameters in the different repetition steps. Since also the other reviewer had problems with this term (he used interactive instead of iterative), it is probably the best choice to avoid this term. We propose to use “repeating spline algorithm” instead of “iterative spline algorithm”.

4. Done, annotation: former figure 7b is now 7c.

5. Yes, fig. 3 points out that besides specific conditions (wavelength ≈ sampling difference) there are no problems. However, in practice, atmospheric profiles / data series are often limited in their altitude or time range. So, it is necessary to set the sampling point differences near to the length of present wave structures. Our approach shows that even in these cases it is possible to get useful results. We replaced the term “surprising”.

6. Both, the intention of this paragraph is to point out that the repeating approach shows the mentioned features for both backgrounds. We tried to make it clearer in the manuscript. We also added to the figure caption “The range of the x- and y-axis differs from the ones used in fig. 3 and 6 (a)”.The scaling of the sum of squared residuals depends on the data resolution since it is the sum over the altitude range 20–40km. The range of the sampling point is extended for the additional wavelengths of 5 and 13 km for the more realistic example in contrast to the pure theoretical examples mentioned before.

7. Improved.

8. Due to the comments of the other reviewer, we included such a comparison already for the CIRA example (in figure 6). Since the main point—the different approximation of traditional and repeating spline—can also be observed in the CIRA example, we hope that it is ok not to show a more or less identical figure here for one SABER profile. Furthermore, we extended the first paragraph of page 9 where we provide possible explanations for the features observed in figure 8.

9. (i) See page 9, first paragraph for possible explanations. The answer to the major comment, number 3 of the other reviewer helps to better understand the information which we added to this paragraph.

(ii) I am not quite sure that I understood every point here correctly but I hope that I can clarify the main points of confusion. Part c and d of figure 8 shows the mean (non-squared) residuals: the mean (non-squared) residuals are calculated as the average over the individual residuals which are indeed much larger and can be positive as well as negative. Due to the latter, we would assume that the mean residual is nearly zero. Obviously, this is not the case: a slight oscillation with an amplitude of ca. 0.5–1 K at 20 km height and ca. 2 K at 100 km height can be observed (in figure 8c and d, logarithmic scaling) for which physical reasons do not exist from our point of view and which is presumably an artefact of the analysis (see page 9, first paragraph for possible explanations).

Smoothing data series by means of cubic splines: quality of approximation and introduction of a~~n~~ ~~iterative~~repeating spline approach

Sabine Wüst [1], Verena Wendt [1,2], Ricarda Linz [1], Michael Bittner [1,3]

[1]Deutsches Fernerkundungsdatenzentrum, Deutsches Zentrum für Luft- und Raumfahrt, 82234 Oberpfaffenhofen, Germany

[2]Umweltforschungsstation Schneefernerhaus, Zugspitze, Germany, now at: Institut für industrielle Informationstechnik, Hochschule Ostwestfalen-Lippe, Ostwestfalen-Lippe, Germany

[3]Institut für Physik, Universität Augsburg, 86159 Augsburg, Germany

Correspondence to: Sabine Wüst (sabine.wuest@dlr.de)

Abstract. Cubic splines with equidistant spline sampling points are a common method in atmospheric science for the approximation of ~~undisturbed~~ background conditions by means of filtering superimposed fluctuations from a data series. What is defined as background or superimposed fluctuation depends on the specific research question. The latter also determines whether the spline or ~~Often, not only the background conditions are of scientific interest but also~~ the residuals—the subtraction of the spline from the original time series—are further analysed.

Based on test data sets, we show that the quality of approximation is not increasing continuously with increasing number of spline sampling points / decreasing distance between two spline sampling points. Splines can generate considerable artificial oscillations in the background and the residuals~~data~~.

We introduce a~~n~~ ~~iterative~~repeating spline approach which is able to significantly reduce this phenomenon. We apply it not only to the test data but also to TIMED-SABER temperature data and choose the distance between two spline sampling points in a way that we are sensitive for a large spectrum of gravity waves.

1 Introduction

It is essential for the analysis of atmospheric wave signatures like gravity waves that these fluctuations are properly separated from the background. Therefore, particular attention must be attributed to this step during data analysis. Splines are a common method in atmospheric science for the approximation of atmospheric background conditions. The shortest wavelength or period which can be resolved by the spline is twice the sampling point distance according to the Nyquist theorem.~~The choice of a sufficiently low number of spline sampling points ensures a smoothing of the original time series. Conclusions about the shortest wavelength / period which is still uniquely resolvable by the spline can be drawn from the application of Shannon's sampling theorem (see e.g. Gubbins, 2004) based on the number of spline sampling points used.~~ Depending on the field of interest, either the smoothed data series or the residuals—the subtraction of a spline from the original time series—are further analysed (see for example the work of Kramer et al., 2016; Baumgarten et al., 2015; Zhang et al., 2012; Wüst and Bittner, 2011; Wüst and Bittner, 2008; Young et al., 1997; Eckermann et al., 1995).

Algorithms for the calculation of splines are implemented in many programming languages and in various code packages making them easy to use. Nevertheless, spline approximations need sometimes to be handled with care when it comes to physical interpretation.

Figure 1 explains our motivation for the work presented below. It shows the squared temperature residuals averaged over one year for the years 2010–2014 versus height between 44° N and 48° N and 5° E~~O~~ and 15° E~~O~~ (approximately 500 profiles per year). This region includes the Alps where gravity waves are supposed to be generated. The vertical temperature profiles are derived from the SABER (Sounding of the Atmosphere using Broadband Emission Radiometry) instrument on board the satellite TIMED (Thermosphere Ionosphere Mesosphere Energetics Dynamics), data version 2.0 (details about this data version can be found in Wüst et al. (2016) and references therein, for example). For the calculation of the residuals, we applied a cubic spline routine with equidistant sampling points. As mentioned above, the shortest wavelength which can be resolved by the spline is two times the distance between two consecutive spline sampling points. At the same time, this wavelength is the largest resolvable one in the residuals. The number of spline sampling points (and the length of the data series) therefore determines the sensitivity of the spline to specific wavelengths. The distance of 10 km between two spline sampling points makes sure that we are sensitive for a large spectrum of gravity waves. Therefore, we take the squared temperature residuals as a simple proxy for gravity wave activity with vertical wavelengths up to 20 km. Obviously, the mean squared residuals do not only reveal a strong and continuous increase with height (note the logarithmic x-axis) as it is expected since gravity wave amplitudes should increase due to the exponentially decreasing atmospheric background pressure with altitude. Superimposed on this general increase of gravity wave activity are well-pronounced oscillations with a wavelength of ca. 10 km, which is nearly equal to the distance between two spline sampling points.

Since we are not aware of any physical reason for this oscillation, we formulate the hypothesis that this is an artefact of the analysis. In order to avoid or at least reduce such problems we here propose an ~~iterative~~repeating variation of the cubic spline approach, which we explain in section 2. In section 3, we apply the original and the ~~iterative~~repeating approach to test data sets. The results are discussed in section 4. A brief summary is given in section 5.

2 Methods and algorithms

The approach we investigate here relies on cubic splines with equidistant sampling points. Since spline theory is well-elaborated, we will therefore not go into much detail here. The algorithm we use is based on Lawson and Hanson (1974).

The first step for the adaption of a spline function to a data series on an interval [a, b] is the choice of the number of spline sampling points (also called knots). These points divide the interval for which the spline is calculated into sub-intervals of equal length. For each sub-interval a third-order polynomial needs to be defined, that means the coefficients have to be determined. At the spline sampling points, not only the function value, but also the first and second derivatives of the two adjacent polynomials need to be equal. The optimal set of coefficients is calculated according to a least square approach where the sum of the squared differences between the data series and the spline is minimized.

As mentioned above, the number of spline sampling points (and the length of the data series) determines the sensitivity of the spline to specific wavelengths. Since the length of the data series must be an integer number of the distance between two spline sample points only certain distances between two consecutive spline sampling points can be chosen if the whole data series is approximated. ~~According to the sampling theorem, the shortest wavelength / period which can be resolved by the spline is two times the distance between two consecutive spline sampling points. At the same time, this wavelength / period is the largest resolvable one in the residuals. Depending on the length of the data series only specific distances between two consecutive spline sampling points can be realised if the whole data series is approximated. The number of spline sampling points (and the length of the data series) therefore determines the sensitivity of the spline to specific wavelengths.~~

~~With SABER data we analysed temperature profiles which extend over a great height distance.~~ We would like to operate the spline algorithm in the way that we provide ~~independently of the ratio between length of the data series and number of spline sampling points~~ the shortest wavelength which shall be resolved by the spline. That means that we have to cut the upper part of the profile from case to case. This is only possible for data sets of sufficient length as the SABER temperature profiles we used for our purpose. In detail, our spline algorithm works as follows. The scheme includes the ~~iterative~~repeating as well as the non-~~iterative~~repeating algorithm.

- *Step 1: Provision of shortest wavelength*

  We provide the algorithm the shortest wavelength which shall be resolved by the spline (in the following denoted as lim). It is equal to the doubled distance between two spline sampling points; therefore the distance between two spline sampling points is equal to lim/2.

- *Step 2: Determination of x-values of the spline sampling points*

The minimal x-value of the data series is subtracted from the maximal x-value, the difference is divided by lim/2. If the result is a whole-number, 1 is added. If this is not the case, the closest integer less than the result is calculated and 1 is added. This is the number of spline sampling points used for the next step. It is denoted as n:

$$n = \left(\frac{x_{max} - x_{min}}{lim/2} - \frac{x_{max} - x_{min}}{lim/2} \bmod 1\right) + 1 \quad (1)$$

Knowing lim/2 and the minimal x-value, the x-values of the further spline sampling points can be calculated.

- *Step 3: Calculation of spline approximation*

  The spline approximation is calculated based on Lawson and Hanson (1974). If the length of the data series is not equal to an integer multiple of lim/2, the surplus part at the end of the data series is not subject of this step. For the non-~~iterative~~repeating approach, the spline algorithm stops here.

- *Step 4 (only in the case of the ~~iterative~~repeating approach): Iteration of starting point*

  The first point of the data series is removed and step 2 and 3 are repeated. If the starting point is equal to the original minimal x-value plus lim/2, the algorithm proceeds with step 5.

- *Step 5 (only in the case of the ~~iterative~~repeating approach): Calculation of the final spline*

  The mean of all splines derived before is calculated. That is the final (~~iterative~~repeating) spline.

For the ~~iterative~~repeating approach, the length of the data series is not the same in each iteration since data at the beginning and the end of the data series are not necessarily part of each iteration: at the beginning of the data series, this holds for all x-values between the minimal x-value and the minimal x-value plus lim/2 (see step 4), at the end of the data series, this is the case for all values between the maximal x-value and the maximal x-value minus lim/2 (see step 3).

For the non-~~iterative~~repeating approach, data are cut only at the end of the data series if the length of the data series is not equal to an integer multiple of lim/2.

3 Case studies

We generate a basic example using an artificial sine with a vertical wavelength of 3 km, a phase of zero and an amplitude of one. The function is sampled every 375 m (that means at its zero-crossings, at its extrema and once in between the zero-crossing and the next extremum / the extremum and the next zero-crossing).

The values for the sampling rate and the vertical wavelength are set arbitrarily. However, the spatial resolution of 375 m is motivated through the spatial resolution of TIMED-SABER, an instrument which is commonly used for the investigation of gravity waves (e.g. Zhang et al., 2012; Ern et al., 2011; Wright et al., 2011; Krebsbach and Preusse, 2007) and which delivered also the temperature profiles we used in fig. 1.

Fig. 2 (a) shows the test data series (dotted line) between 15 km and 100 km height. This great height range is chosen since it facilitates the demonstration of our results. A (non-~~iterative~~repeating) spline with a distance of 1.5 km between two spline sampling points is fitted (solid line). According to the ~~sampling~~ Nyquist theorem, the chosen distance between two spline sampling points is small enough for resolving the oscillation in our test data. In part (b) and (c) of fig. 2, a spline with a distance of 1.6 km and 1.4 km between two spline sampling points is calculated. Part (d) to (f) of fig. 2 focus on the height range of 15 km to 50 km of fig. 2 (a) to (c): here, the height-coordinates of the spline sampling points are plotted additionally (dashed-dotted lines). The asterisks mark the sampling points of the original sine. The spline adaption in fig. 2 (a) / (d) differs significantly from the spline adaption in fig. 2 (b) / (e) and 2 (c) / (f): apart from a slight oscillation at the beginning / end of the height interval, the spline is equal to zero in fig. 2 (a) / (d). The spline approximation plotted in fig. 2 (b) and (c) shows a beat-like structure in the whole height range.

In order to give an overview concerning the quality of adaption not only for some chosen examples as they were shown in fig. 2, the test data set is approximated by a cubic spline with varying numbers of spline sampling points. The squared differences between the spline and the test data are summed up between 20 km and 40 km (this height interval is chosen in order to be consistent with fig. 7 later). We call this value the sum of squared residuals which is equal to the approximation error in this case. It is not decreasing continuously with increasing number of spline sampling points / decreasing distance between two spline sampling points but it is characterized through a superimposed oscillation which reaches its maximum for a distance of ca. 1.5 km between two spline sampling points (fig. 3, solid line). When changing the phase of the test data set to $\pi/2$ (instead of zero), the sum of squared residuals~~approximation error~~ for a distance of ca. 1.5 km between two spline sampling points is much lower (fig. 3, dashed line), that means the sinusoidal oscillation is better adapted in this case. This makes clear that the sum of squared residuals~~approximation error~~ depends on the phase of the oscillation (one can also say on the exact position of the spline sampling points).

The analysis described above is repeated, but the phase of the oscillation is varied between 0 and 2π. The sum of squared residuals~~approximation error~~ (between 20 km and 40 km) is calculated for three different distances between two spline sampling points: 1.5 km (fig. 4, solid line), 1.4 km (fig. 4, short dashes) and 1.6 km (fig. 4, long dashes). The dependence on the phase is most pronounced for a distance of 1.5 km: the oscillation is adapted very well between 20 km and 40 km for a phase of π /2 and 3π/2. For a phase of 0 and π, the contrary holds.

This example directly motivates the application of the ~~iterative~~repeating spline approach on the same test data set (see fig 5 (a)–(f) which can be directly compared to fig 2 (a)–(f): the black line represents the final spline approximation and the different colours refer to the spline approximations during the different iteration steps). In this case, the sum of squared residuals~~approximation error~~ depends much less on the distance between two spline sampling points (fig. 6 (a)) and on the phase of the test data set (fig. 6 (b)). Only for a distance of 1.6 km between two spline sampling points, a slight phase dependence is still visible (fig. 6 (b)).

Until now, we showed only test data which are not superimposed on a larger-scale variation like the atmospheric temperature background. Now, three sinusoidals with vertical wavelengths of 3, 5, and 13 km, phase 0, π/3, and π/5, and amplitude ~~0.5~~2.0 (growing amplitude with height neglected for simplicity reasons) are superimposed on a realistic vertical temperature background (fig. 7 (a)). The background is based on CIRA-86 (COSPAR International Reference Atmosphere, Committee on space Research and NASA National Space Science Data Center, 2006) temperature data for 45° N for January, which is brought on a regular grid using a cubic spline with a distance of 3 km between two spline sampling points. It was checked that no additional signatures are caused thereby. The sum of squared residuals ~~approximation error~~ shows three steps but no superimposed oscillations (fig. 7 (b)): the first step at ca. 6 km to 7 km (distance between two spline sampling points), the second one at ca. 2 km to 3 km and the last one at 1 km to 2 km. Following ~~Shannon~~Nyquist's sampling theorem, this observation can be explained through the ability of the spline to adapt the original wavelengths. The realistic background makes clear why we restrict the calculation of the sum of squared residuals~~approximation error~~ to the height range between 20 km and 40 km: this height interval is chosen in order to exclude especially the stratopause since the fast changing temperature gradient can cause additional problems for the spline approximation. Furthermore, the choice of this interval makes sure that the data used for fig. 7 (b) are part of each iteration step (which is not the case for the data at the beginning and the end of the data series, see section 2).

4 Discussion

In section 3, we showed that the ability of a spline to approximate oscillations and therefore to filter for a specific part of the wave spectrum varies

a) with the number of spline sampling points, and
b) with the exact position (height coordinate) of the spline sampling points.

While the first statement can be explained ~~Shannon~~Nyquist's sampling theorem, the second one is not well-known~~might be surprising~~.

When the distance between two spline sampling points matches exactly half the wavelength of the test data, the approximation is worst for a phase of 0 and π. In this case, the spline sampling points are located exactly between the extrema of the test data. If the height coordinates of the spline sampling points agree with the height coordinates of the extrema of the test data, the contrary holds (in appendix A, we provide a mathematical explanation for this observation). The dependence of the quality of approximation on the phase of the test data decreases with greater / smaller distances between two spline sampling points (fig. 3). These findings directly motivate the use of the presented ~~iterative~~repeating spline approach which is characterized by varying positions (height coordinate) of the spline sampling points.

Furthermore, we showed that if the distance between two spline sampling points is only slightly larger or smaller than half the wavelength present in the data series and if enough wave trains are present (which might not be the case in reality), the (non-~~iterative~~repeating) spline reminds of a beat (see fig. 2 (b) and (c), an explanation is given in appendix B). The subtraction of such a "beat" will lead to an artificial oscillation in the residuals with a periodically increasing and decreasing amplitude reaching ca. 70–80% of the original amplitude at maximum (fig. 2 (e) and (f)). This oscillation must not be interpreted as a gravity wave of varying amplitude, for example, and the described effect has to be taken into account when analysing wavelengths similar to the doubled distance between two spline sampling points.

For our case studies, we used a constant and a realistic CIRA-based temperature background profile. For both background profiles, ~~W~~we showed the sum of squared residuals~~approximation error (the squared differences between the spline and the test data summed up between 20 km and 40 km)~~ decreases much smoother with increasing number of spline sampling points for the ~~iterative~~repeating approach compared to the non-~~iterative~~repeating one (compare fig. 3 to fig. 6 (a)) and the amplitude of the "beat"-like structure is reduced.

However, the motivation for this work was—as already mentioned—the results shown in fig. 1 which are characterized by a strong superimposed oscillation with a wavelength of approximately 10 km for which we do not have a physical explanation. Figure 8 (a) now depicts the mean squared residuals after the application of the ~~iterative~~repeating spline to the same data set, fig. 8 (b) focuses on the year 2014 (the dashed line is based on the application of the ~~iterative~~repeating spline, the solid line refers to the non-~~iterative~~repeating spline). This year is chosen arbitrarily and allows the direct comparison between the ~~iterative~~repeating and non-~~iterative~~repeating approach. The amplitude of the superimposed oscillation is reduced significantly but the oscillation can still be observed. This supports our hypothesis that the strong superimposed oscillation described in fig. 1 is an artefact of the non-~~iterative~~repeating spline de-trending procedure. Furthermore, it becomes obvious now that gravity wave activity increases less with altitude~~is less variable~~ between approximately 45 km and 60 km height compared to the height range below and above. This is in accordance with literature (e.g. Mzé et al., 2014; Offermann et al., 2009). For most heights, the mean squared residuals are smaller for the ~~iterative~~repeating approach than for the non-~~iterative~~repeating one. At 38 km height for example, the difference reaches ca. 2.5 K², which is approximately 32 % (referring to the mean value of both approaches). Shuai et al. (2014) use an earlier version of TIMED-SABER temperature data (1.07) and a different de-trending procedure as we do in order to derive monthly averages of the squared temperature fluctuations for the years 2002–2010. They provide this parameter in dB ($10 \cdot log_{10}\,(T'^{2}_{GW})$) with the squared temperature fluctuation T'^{2}_{GW} which makes a short calculation necessary.

For 100 km height, we extract a yearly mean of ca. 21 dB for 50°N from their figure 2 that means

$$10 \cdot log_{10}\,(T'^{2}_{GW}) = 21 \;\Leftrightarrow log_{10}\,(T'^{2}_{GW}) = 2.1 \Leftrightarrow T'^{2}_{GW} = 10^{2.1} \approx 126$$

For 25 km height, we read am mean value of ca. 4 dB:

$$10 \cdot log_{10}\,(T'^{2}_{GW}) = 4 \;\Leftrightarrow log_{10}\,(T'^{2}_{GW}) = 0.4 \Leftrightarrow T'^{2}_{GW} = 10^{0.4} \approx 2.5$$

These values agree very well with the ones provided here. Furthermore, the overall structure which is characterized by a slow increasing or even a nearly constant gravity wave activity in the upper stratosphere can be observed in their fig. 2 and our fig. 8 (a).

In order to give a comprehensive comparison of the ~~iterative~~repeating and the non-~~iterative~~repeating spline algorithm, we calculate also the mean (non-squared) residuals. In this case, the results look very similarly; in both cases, they again show an oscillation with a vertical wavelength of 10−20 km (fig. 8 (c) for the non-~~iterative~~repeating approach, fig. 8 (d) for the ~~iterative~~repeating spline approach). We can explain this in the following way: when calculating the mean (non-squared) residuals and the mean squared residuals at a specific height, one refers to two different parameters of the distribution of residuals at that specific height. While the mean (non-squared) residuals estimate the mean of the distribution, the mean squared residuals refer to the variance of the distribution. We conclude: at a defined height, the ~~iterative~~repeating approach changes the mean of the distribution of the residuals only slightly, but it reduces its spread significantly. For individual profiles, the approximation through the ~~iterative~~repeating approach is therefore less variable in average and can be recommended. The ~~iterative~~repeating approach can also be recommended if squared residuals are needed for further analysis (e.g. for the calculation of the wave potential energy). If non-squared residuals will be analysed, it does not make a difference in average which approach is applied. In this case, only waves with amplitudes larger than 0.5 K in the stratosphere and 1.0 K in the mesosphere (fig. 8 (c) and (d)) should be taken seriously.

It is known that the tropo-, strato-, and mesopause ~~height regions~~ where the temperature gradient changes are challenging for approximation methods. The same holds for the beginning and the end of a data series. Since the beginning and the end of the data series do not very much (we analysed the SABER data between ca. 12–14 km and 110 km height), possible side effects might not cancel out.~~,~~ ~~and w~~We speculate that these~~is~~ ~~is~~ are at least ~~one~~two of the reasons for the superimposed oscillation in fig. 8 (c) and (d). However, analysing this hypothesis is beyond the scope of this manuscript.

There exist many methods to approximate / de-trend / filter time series (see e.g Baumgarten et al. (2015), and references therein) and we do not claim that the presented ~~iterative~~repeating cubic spline is the best method for every purpose and every data series. It is just one possible algorithm which reduces disadvantages of the non-~~iterative~~repeating cubic spline routine as it was proposed by Lawson and Hanson (1974) like the dependence of approximation on the exact position (height coordinate) of the spline sampling points. However, it comes along with enhanced computational effort which is of special importance when analysing large data sets.

5 Summary

It is essential for the analysis of atmospheric wave signatures like gravity waves that these fluctuations are properly separated from the background. Therefore, particular attention must be attributed to this step.

Cubic splines with equidistant sampling points are a common method in atmospheric science for the approximation of superimposed, large-scale structures in data series. The subtraction of the spline from the original time series allows the investigation of the residuals by means of different spectral analysis techniques. However, splines can generate artificial oscillations in the residuals. The ability of a spline to approximate oscillations varies not only with the number of spline sampling points, but also with their exact position. When the distance between two spline sampling points equals exactly or approximately half the wavelength of the waves present in the data, the last-mentioned effect is most pronounced.

Since knowledge about the wavelengths present in the data set is normally not available in advance, this directly motivates the use of an ~~iterative~~repeating spline which is based on changing starting points. It comes along with enhanced computational effort but it can be recommended for the approximation / de-trending of individual profiles and if squared residuals are needed for further analysis (e.g. for the calculation of the wave potential energy).

Acknowledgement

We would like to thank the TIMED-SABER team for their great work in providing an excellent data set.
We also thank the Bavarian Ministry for Environment and Consumer Protection for financially supporting our work: V. Wendt was paid by the Bavarian project BHEA (Project number TLK01U-49580, 2010–2013). The work of S. Wüst was subsidised in parts by this project.
At least, we thank Julian Schmoeckel, University of Augsburg, for helping to produce the test data sets and the figures.

Appendix A

Between two spline sampling points, a spline is equal to a cubic polynomial of the form

$$f(z) = az^3 + bz^2 + cz + d \qquad \text{with} \qquad a, b, c, d \in \mathbb{R}. \tag{2}$$

Its derivatives are

$$f^{(1)}(z) = 3az^2 + 2bz + c, \tag{3}$$

$$f^{(2)}(z) = 6az + 2b, \tag{4}$$

$$f^{(3)}(z) = 6a. \tag{5}$$

Between two spline sampling points, the second derivative of a spline depends linearly on the height coordinate z. That means the curvature of the spline can change from negative to positive or vice versa between two spline sampling points but it can only increase or decrease linearly or it can stay constant. At the spline sampling points, all derivatives of the two adjacent polynomials must agree. For example, a spline cannot form two parabolas with different signs in two adjacent intervals in order to approximate a sine / cosine since the second derivative (curvature) would be positive constant in one interval and negative constant in the other. If the spline sampling points are not distributed in a way such that the curvature of the original function increases or decreases linearly between two spline sampling points, the spline cannot approximate the original function properly.

Therefore, the ability of the spline to reproduce a sine / cosine does not only depend on the number of spline sampling points, it varies also with their position.

Appendix B

The optimal spline parameters are determined through a least-square approach: depending on the spline parameters, the squared differences between the spline and the original data set are minimized. The maximum wavelength which a spline can approximate in principle is equal to two times the distance between two spline sampling points.

Let us denote the oscillation which has to be approximated with $f_1(z)$ and the spline with $f_2(z)$.

If those two oscillations which will be subtracted from each other are characterized by very similar wave numbers $k_1$and k_2, then a beat with the following wave numbers will occur.

$$f_1(z) - f_2(z) = \sin k_1 z - \sin k_2 z \overset{(*)}{=} 2\cos\left(\frac{k_1+k_2}{2}z\right)\sin\left(\frac{k_1-k_2}{2}z\right) \tag{6}$$

where

$\frac{k_1+k_2}{2}$ is the wave number of the beat, which is very similar to the original wave number, and

$\frac{k_1-k_2}{2}$ is the wave number of the envelope.

$(*)$: application of an addition theorem

References

Baumgarten, G., Fiedler, J., Hildebrand, J., and Lübken, F.-J.: Inertia gravity wave in the stratosphere and mesosphere observed by Doppler wind and temperature lidar, Geophys. Res. Lett., 42, 10,929–10,936, doi: 10.1002/2015GL066991, 2015.

Committee on Space Research; NASA National Space Science Data Center: COSPAR International Reference Atmosphere (CIRA-86): Global Climatology of Atmospheric Parameters. NCAS British Atmospheric Data Centre. http://catalogue.ceda.ac.uk/uuid/4996e5b2f53ce0b1f2072adadaeda262, 2006.

Eckermann, S. D., Hirota, I., and Hocking, W. K.: Gravity wave and equatorial wave morphology of the stratosphere derived from long-term rocket soundings. Q.J.R. Meteorol. Soc., 121, 149–186, doi: 10.1002/qj.49712152108, 1995.

Ern, M., Preusse, P., Gille, J. C., Hepplewhite, C. L., Mlynczak, M. G., Russell III, J. M., and Riese, M.: Implications for atmospheric dynamics derived from global observations of gravity wave momentum flux in stratosphere and mesosphere. J. Geophys. Res., 116, doi. 10.1029/2011JD015821, 2011.

~~Gubbins, D.: Time series analysis and inverse theory for geophysicists, Cambridge University Press, 2004.~~

Kramer, R., Wüst, S., and Bittner, M.: Climatology of convectively generated gravity waves at Prague based on operational radiosonde data from 13 years (1997–2009). J. Atmos. Sol.-Terr. Phys., 140, 23–33, doi: 10.1016/j.jastp.2016.01.014, 2016.

Krebsbach, M., and Preusse, P.: Spectral analysis of gravity wave activity in SABER temperature data, Geophys. Res. Lett., 34, L03814, doi: 10.1029/2006GL028040, 2007.

Lawson, C. L., and Hanson, R. J.: Solving least squares problems, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, U.S.A. 1974.

Mzé, N., Hauchecorne, A., Keckhut, P., and Thétis, M.: Vertical distribution of gravity wave potential energy from long-term Rayleigh lidar data at a northern middle-latitude site, J. Geophys. Res. Atmos., 119, 12,069–12,083, doi: 10.1002/2014JD022035, 2014.

Offermann, D., Gusev, O., Donner, M., Forbes, J. M., Hagan, M., Mlynczak, M. G., Oberheide, J., Preusse, P., Schmidt, H. and Russell III, J. M.: Relative intensities of middle atmosphere waves, J. Geophys. Res., 114, D06110, doi: 10.1029/2008JD010662, 2009.

Shuai, J., Zhang, S., Huang, C., YI, F., Huang, K., Gan, Q., and Gong, Y.: Climatology of global gravity wave activity and dissipation revealed by SABER/TIMED temperature observations. Science China Technological Sciences, 57, 998–1009, doi: 10.1007/s11431-014-5527-z, 2014.

Wright, C. J., Rivas, M. B., and Gille, J. C.: Intercomparisons of HIRDLS, COSMIC and SABER for the detection of stratospheric gravity waves, Atmos. Meas. Tech., 4, 1581-1591, doi: 10.5194/amt-4-1581-2011, 2011.

Wüst, S., and Bittner, M.: Gravity wave reflection: case study based on rocket data. J. Atmos. Sol.-Terr. Phys., 70, 742–755, doi:10.1016/j.jastp.2007.10.010, 2008.

Wüst, S., and Bittner, M.: Resonant interaction between two planetary waves with zonal wave number two? J. Atmos. Sol.-Terr. Phys., 73, 771–778, doi:10.1016/j.jastp.2011.01.004, 2011.

Wüst, S., Wendt, V., Schmidt, C., Lichtenstern, S., Bittner, M., Yee, J.-H., Mlynczak, M. G., and Russell III, J. M.: Derivation of gravity wave potential energy density from NDMC measurements. J. Atmos. Sol.-Terr. Phys., 138–139, 32-46, doi:10.1016/j.jastp.2015.12.003, 2016.

Young, L. A., Yelle, R. V., Young, R., Seiff, A., and Kirk, D. B.: Gravity Waves in Jupiter's Thermosphere, Science, 276, 108–111, doi:10.1126/science.276.5309.108, 1997.

Zhang, Y., Xiong, J., Liu, L., and Wan, W.: A global morphology of gravity wave activity in the stratosphere revealed by the 8-year SABER/TIMED data, J. Geophys. Res., 117, D21101, doi:10.1029/2012JD017676, 2012.

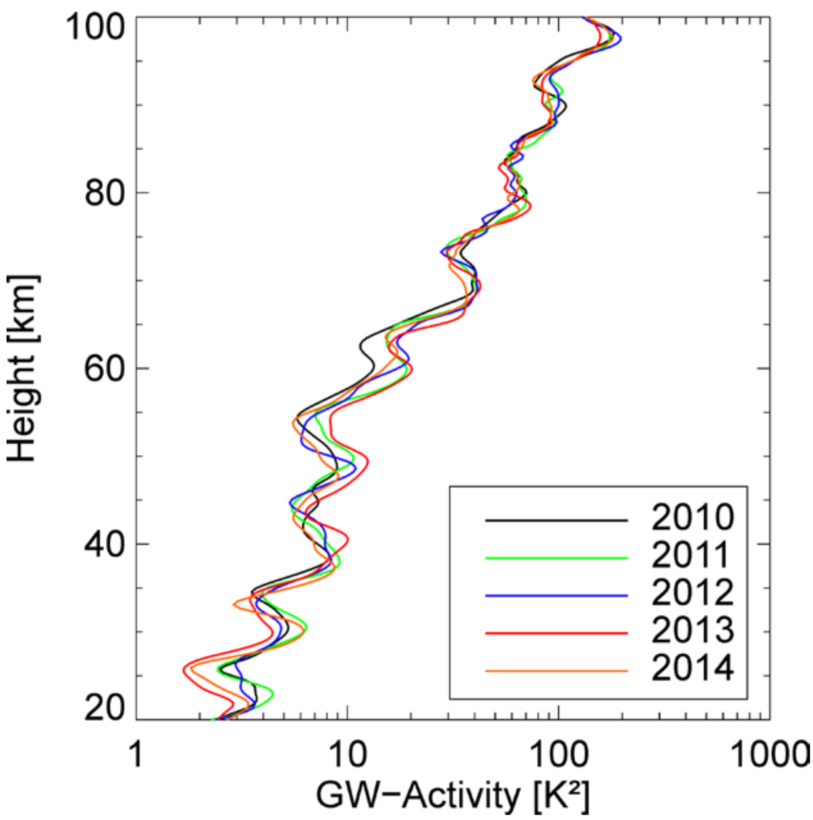
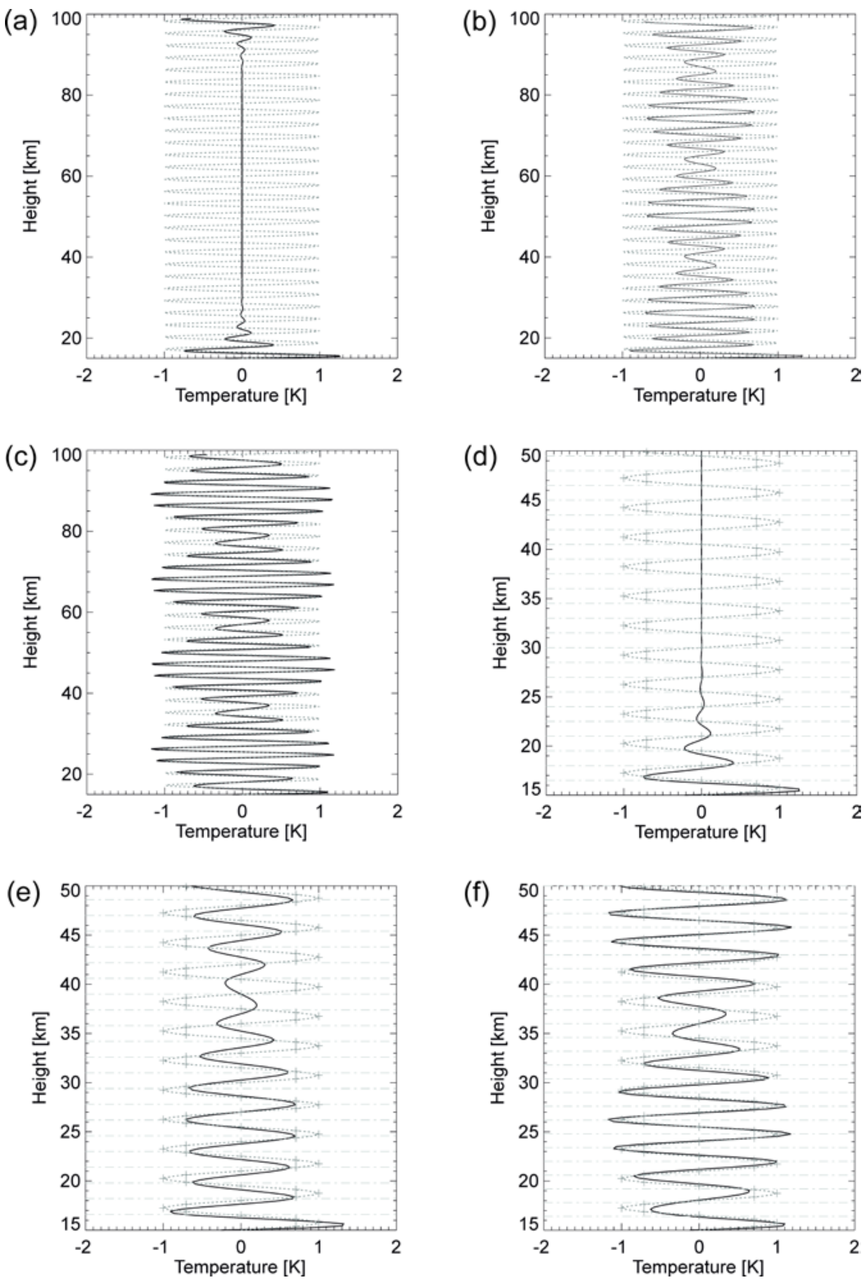


Figure 1: Mean squared temperature residuals for the years 2010 to 2014 (colour-coded): they are derived from TIMED-SABER, data version 2.0 by using a cubic spline routine with equidistant sampling points for de-trending. The distance between two spline sampling points is 10 km. All vertical SABER temperature profiles which were retrieved between 44° N and 48° N and 5° O and 15° O are used (that means approximately 30-50 profiles per month and approximately 500 profiles per year).



(a)
(b)
(c)
(d)
(e)
(f)
Height [km]
Temperature [K]

Figure 2: This figure shows the approximation of a cubic spline using different numbers of spline sampling points:

(a) A spline with a distance of 1.5 km between two spline sampling points is fitted (solid line) to the test data (dotted line).

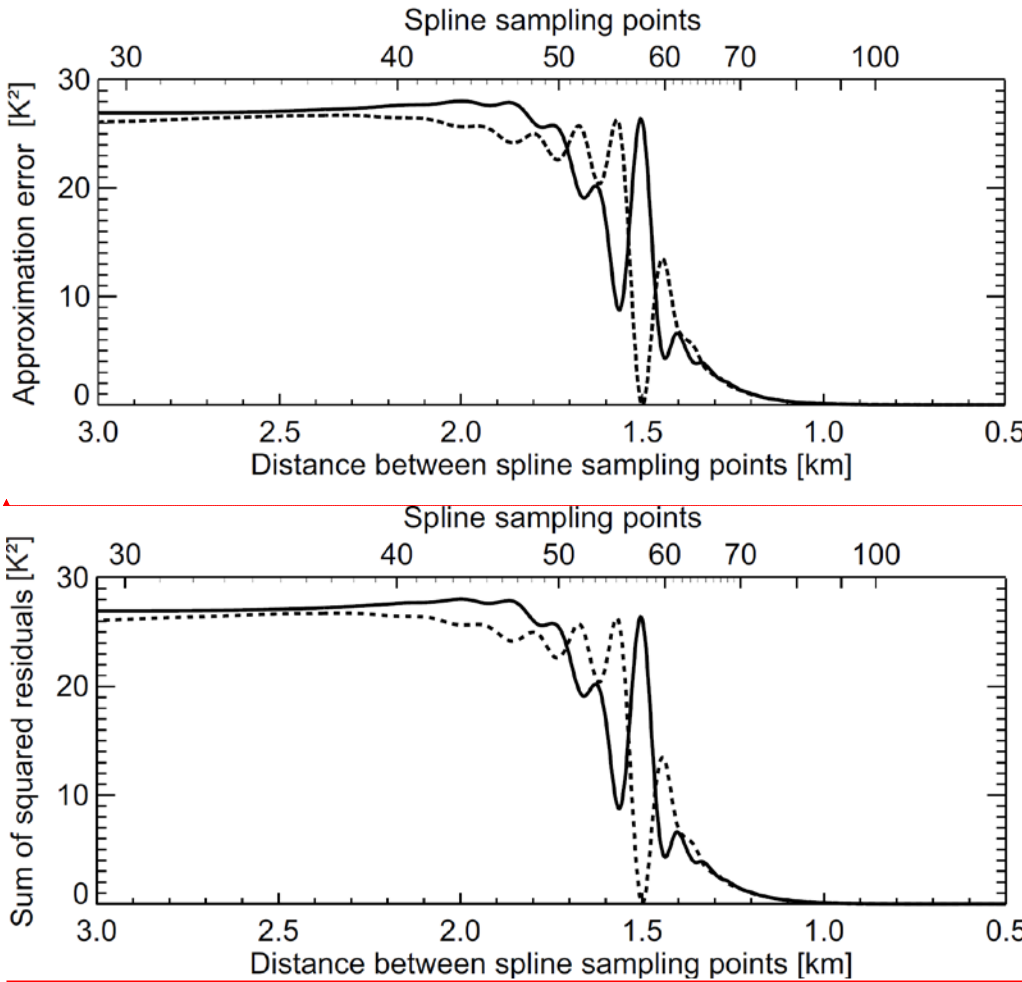
(b) Same as (a) but the distance between two spline sampling points is 1.6 km.

(c) Same as (a) but the distance between two spline sampling points is 1.4 km.

(d) Same as (a) but restricted to the height range between 15 km and 50 km. The dashed-dotted lines refer to the height-coordinate of the spline sampling points. The asterisks show the sampling points of to the original sine.

(e) Same as (b) but restricted to the height range between 15 km and 50 km.

(f) Same as (c) but restricted to the height range between 15 km and 50 km.



Formatiert: Englisch (Großbritannien)

Figure 3: This figure shows the differences between the spline and the approximated test data (solid line: phase of 0, dashed line: phase of π/2) which are summed up between 20 km and 40 km. They are plotted versus the distance between the number of spline sampling points / distance between spline sampling points. The number of spline sampling points / distance between spline sampling points refers to the whole height range between 15 km and 100 km.

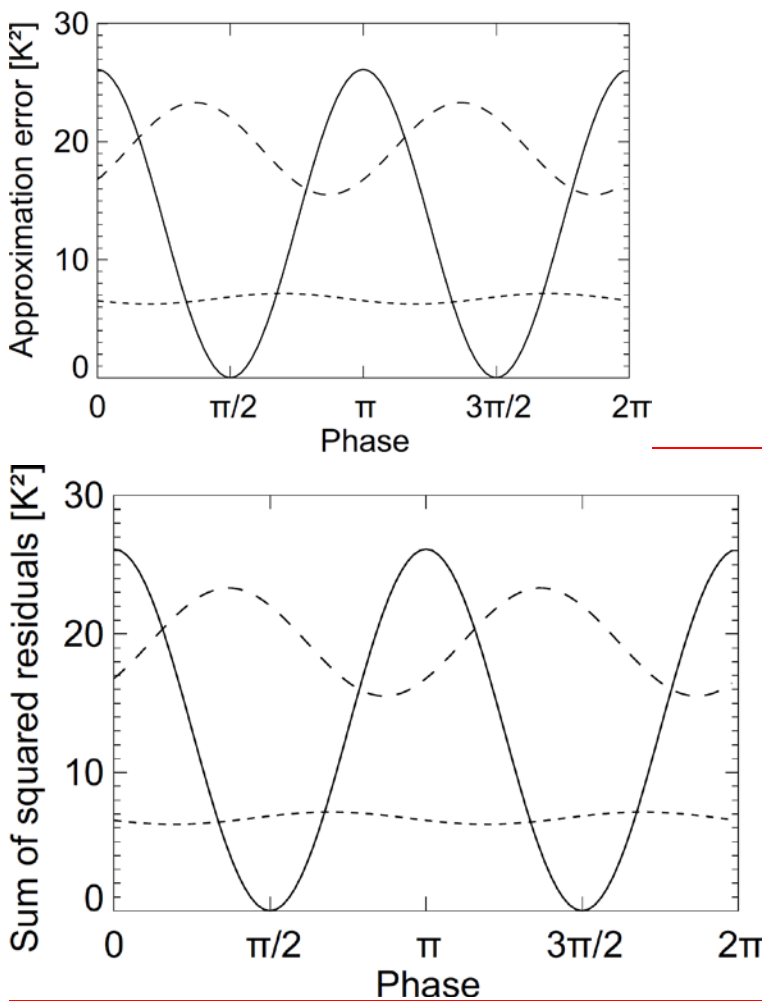
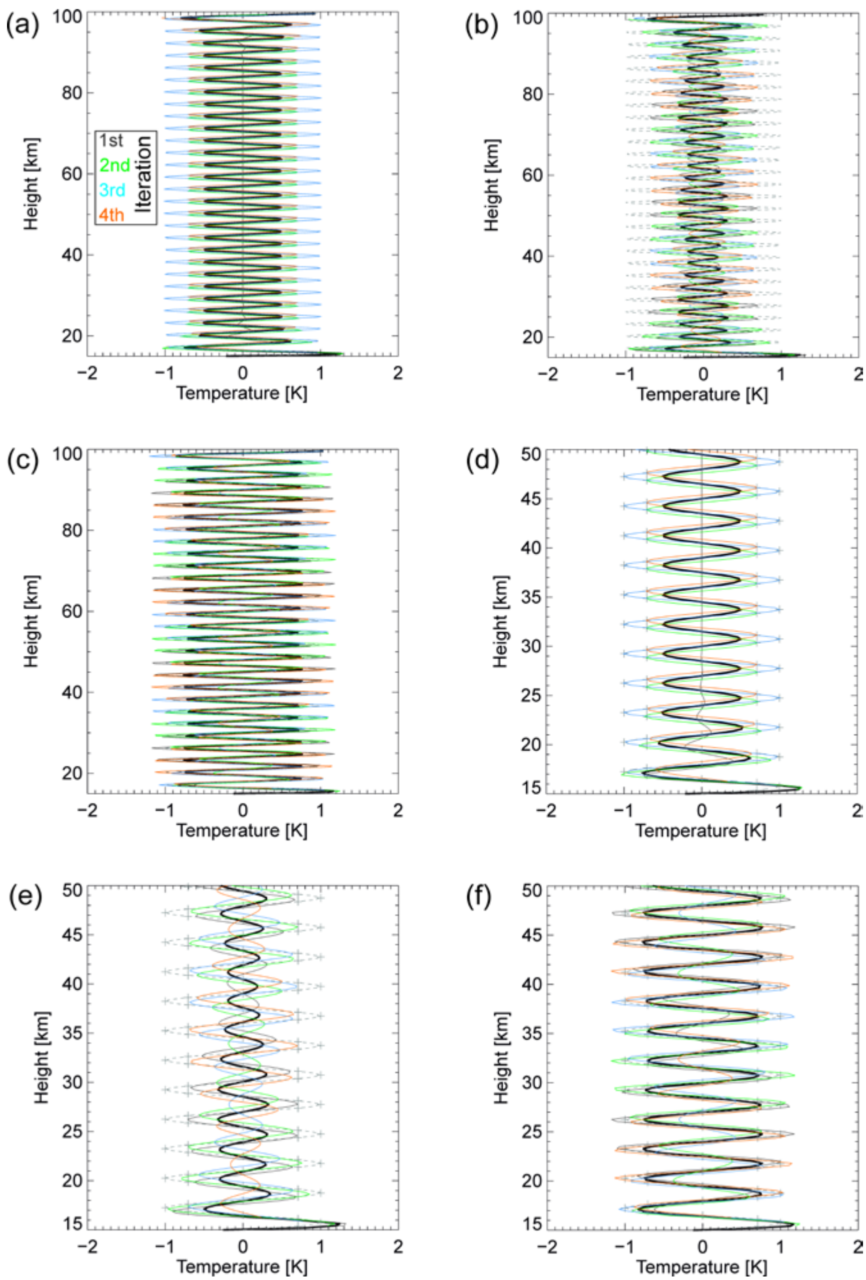


Figure 4: Dependence of the sum of squared residuals ~~approximation error~~ on the phase of the wave with a wavelength of 3.0 km and a distance of 1.4 km (short dashes), 1.5 km (solid line) and 1.6 km (long dashes) between two spline sampling points.

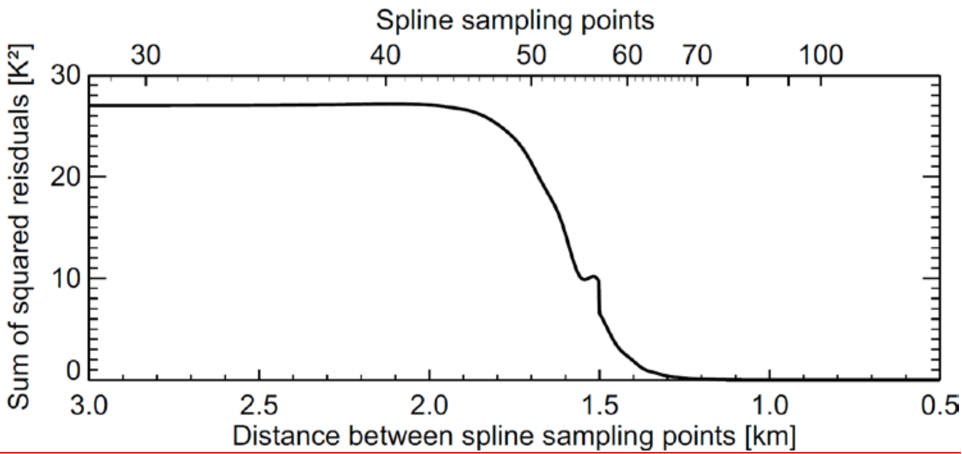
Formatiert: Englisch (Großbritannien)

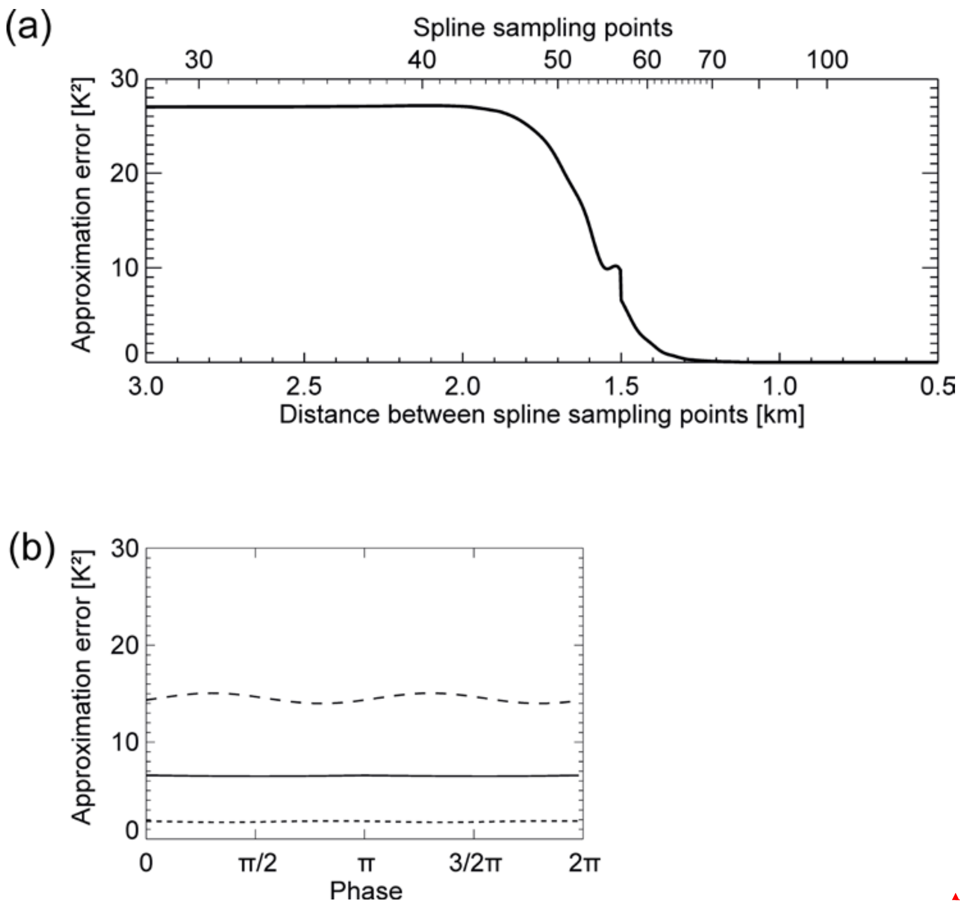


(a)
(b)
(c)
(d)
(e)
(f)
Height [km]
Temperature [K]
1st
2nd
3rd
4th
Iteration

Figure 5: Here, the results based on the ~~iterative~~repeating spline approach are shown. The different colours refer to the different spline approximations (to keep it as clear as possible, we only show the first four iterations, a fifth one exists for case (b) and (e), see step 4 of the algorithm). The black line represents the final spline approximation. The distance between two spline sampling points in part (a) to (f) agrees with the respective values in figure 2 part (a) to (f). While part (a) to (c) show the height range between 15 km and 100 km, part (d) to (f) focus on the height range between 15 km to 50 km. The asterisks ~~and dashed-dotted lines~~ have the same meaning as in figure 2

(a)


Spline sampling points
30
40
50
60
70
100
Sum of squared reisduals [K²]
30
20
10
0
3.0
2.5
2.0
1.5
1.0
0.5
Distance between spline sampling points [km]



Formatiert: Englisch (Großbritannien)

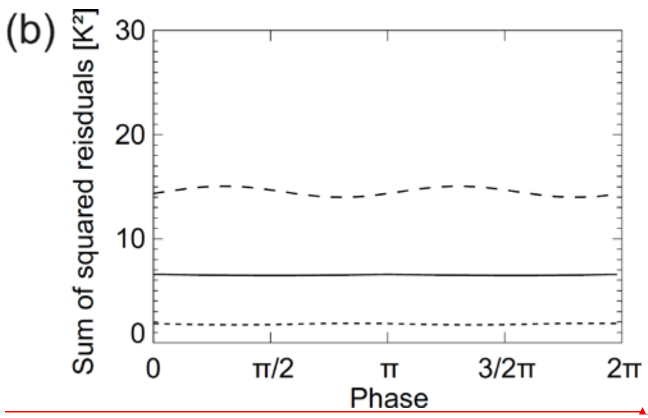
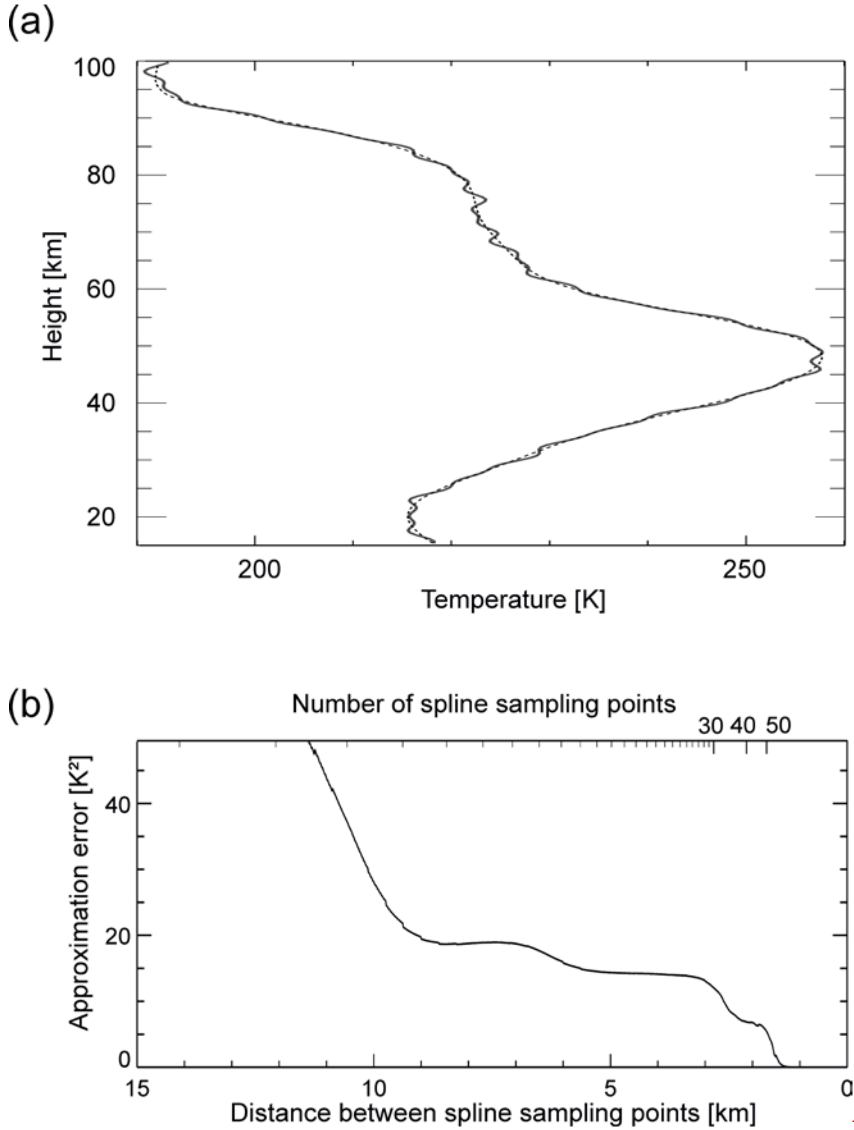


Figure 6: Part (a) is equivalent to figure 3, part (b) is equivalent to figure 4, but here the ~~iterative~~repeating spline approach is used.

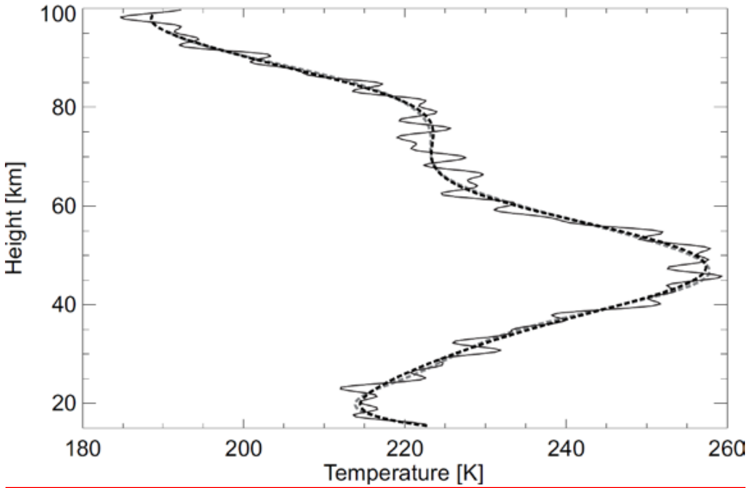
Formatiert: Englisch (Großbritannien)

Formatiert: Englisch (Großbritannien)

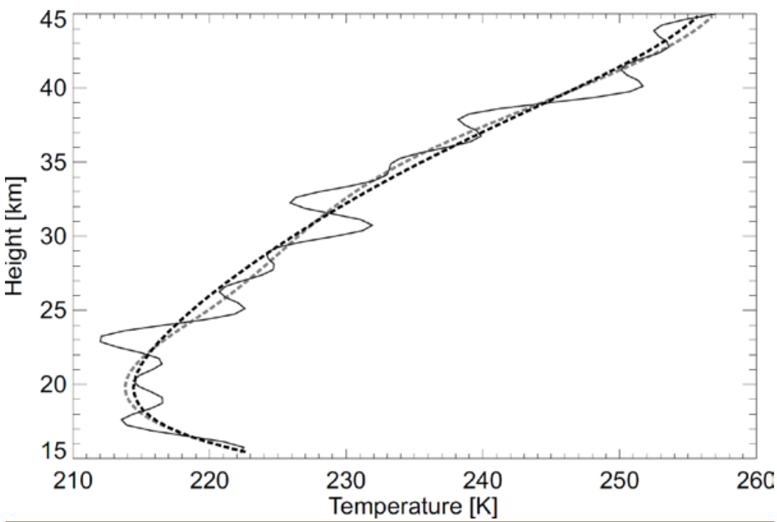
(a)



(a)
100
80
60
40
20
Height [km]
200
250
Temperature [K]
(b)
Number of spline sampling points
30 40 50
40
20
0
Approximation error [K²]
15
10
5
0
Distance between spline sampling points [km]


Height [km]
100
80
60
40
20
180
200
220
240
260
Temperature [K]

(b)


Height [km]
45
40
35
30
25
20
15
210
220
230
240
250
260
Temperature [K]

(c)

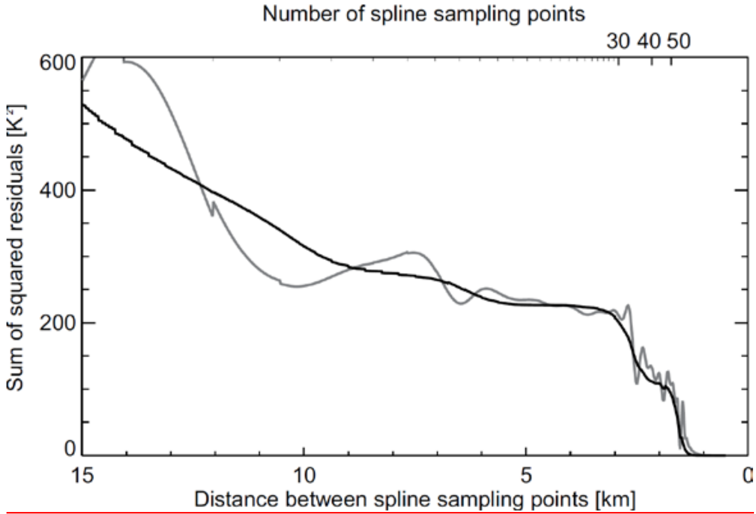


Figure 7: a) The solid line depicts three sinusoidals with vertical wavelengths of 3, 5 and 13 km, phase 0, π/3, and π /5, and amplitude ~~0.5~~2.0 ~~which are superimposed on a realistic temperature profile~~and the spline approximations for the repeating (black) and non-repeating (grey) spline approach (dashed line, sampling point distance 10 km). Part b) as part a) but focussing on the height interval between 20 and 40 km for which the sum of squared residuals is calculated in part c). The range of the x- and y-axis differs from the ones used in fig. 3 and 6 (a).~~Part b) shows the approximation error depending on the distance between two spline sampling points.~~

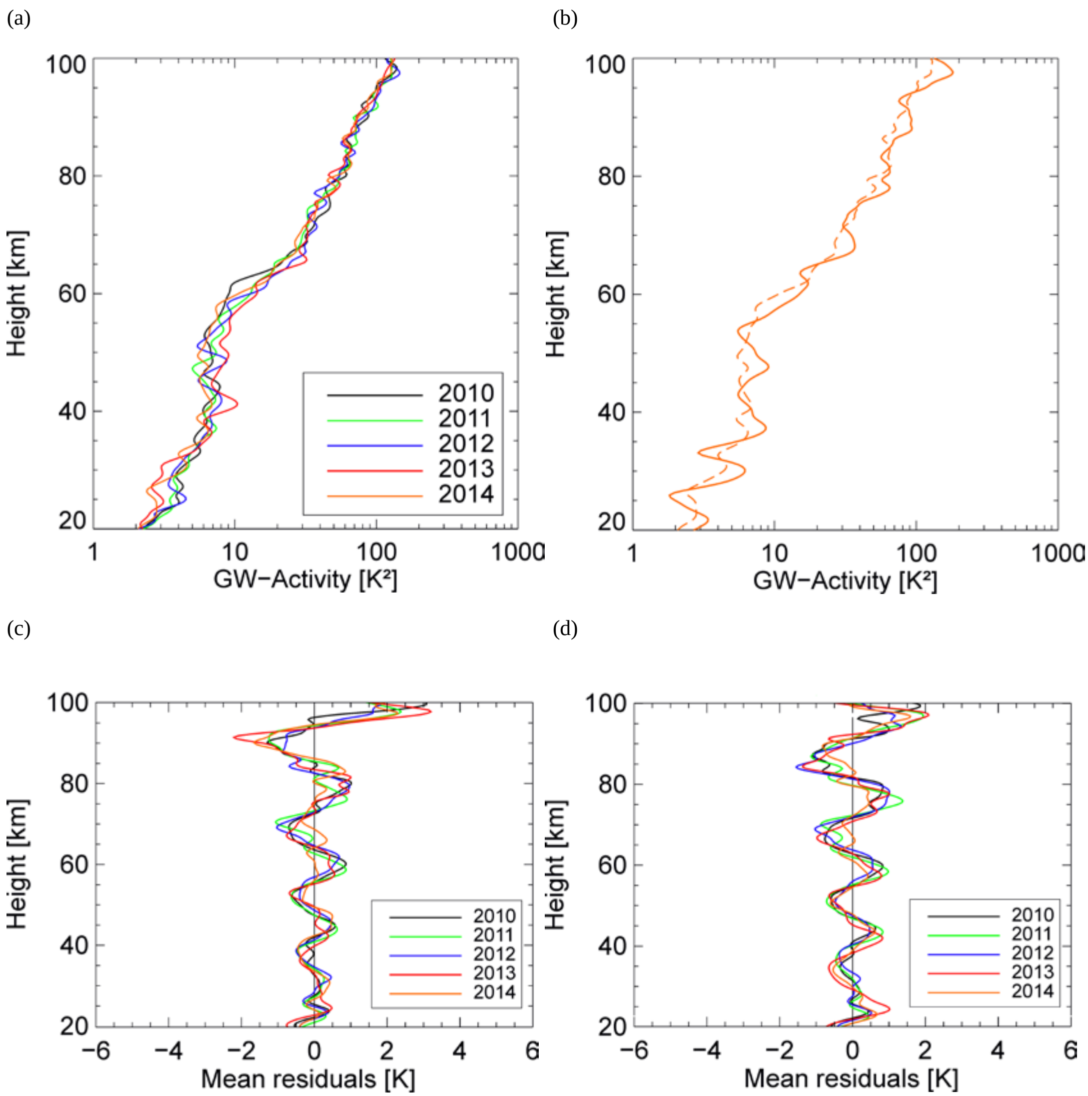


Figure 8: a) As figure 1, for de-trending the ~~iterative~~repeating cubic spline routine with equidistant sampling points as it is described in section 2 is used. b) Mean squared residuals for the ~~iterative~~repeating (dashed line) and the non-~~iterative~~repeating (solid line) approach for the year 2014. c) Mean (non-squared) residuals for the *non-~~iterative~~repeating* approach for the years 2010 to 2014, and d) mean (non-squared) residuals for the *~~iterative~~repeating* approach.