Review of paper submitted to Ocean Science: “On the resolutions of ocean altimetry maps” by Ballarotta, Ubelmann, Pujol, Taburet, Fournier, Legais, Faugere, Chelton, Dibarboure, and Picot

J. Thomas Farrar
Department of Physical Oceanography
Woods Hole Oceanographic Institution

March 27, 2019

This paper presents a new method for assessing the spatial and temporal resolution of a gridded data product and applies the method to the DUACS altimetry product. The paper is well written and the undertaking is worthwhile. The method for assessing resolution is based on estimation of the coherence between the input data (or, in some cases, independent data) and the gridded product. I have a great deal of respect for these authors, and so it is with a lot of discomfort and hesitation that I must say that the method used in the paper is fundamentally flawed, and I strongly recommend against publication. Unfortunately, I do not believe there is any way to salvage the paper. In what follows, I will focus on what is wrong with the proposed method of assessing resolution, because that is essentially the only problem I have with the paper. I would like to express my sincere regret to the authors that I must argue so strongly against publication of their paper– I hope they find this review helpful rather than offensive. I also apologize to all involved for the lengthy review! I felt it was necessary to include some examples to illustrate what is wrong with the methodology. If my assessment is incorrect, I would be happy to be corrected.

The method is based on a conceptual model of a linear single-input/single-output system (in the terminology of the book by Bendat and Piersol that the authors cite) that is used to interpret the coherence of the data product with other data (such as the input data). The conceptual linear system has no noise on the measurement of the system output but does have noise on the measurement of the system input. As described by the authors and by Bendat and Piersol (2010, p. 185), there is an input signal u(t) that goes through a linear system (the mapping) to produce an output signal y(t). The output y(t) is known perfectly (without measurement noise), but the input u(t) is not known perfectly– only a noisy measurement of it is available, x(t)=u(t)+m(t). For brevity, I will refer to this as a ”Case 2” model for interpretation of the coherence.

The fundamental problem is that the proposed measure of resolution does not take into account the filtering by the mapping system. The estimated coherence is affected by this filtering, but not in a way that makes the coherence a useful measure of resolution. One problem is that the coherence in space is affected by the smoothing in time and vice versa. In addition, even without this problem (e.g. in a one dimensional system), the coherence reveals nothing useful about the smoothing in the mapping system. I will provide some examples to demonstrate both of these points.

Before showing some examples to illustrate the problem, I would like to note the following things that bother me about the method:

1. A first objection is that this model of the mapping system is not appropriate. The input to the mapping system is a signal plus noise (i.e., x(t)=u(t)+m(t)). That is, the noisy along-track data are the input to the mapping system. The Bendat-Piersol scenario is for a case where we have imperfect knowledge of the inputs to a linear system and perfect knowledge of the output. In the case of the DUACS system, we know the inputs and outputs perfectly (to within roundoff errors of machine precision). The only way the Case 2 model could apply here is if we are willing to assume that the mapping system eliminates all of the measurement errors and sampling errors (like aliasing) and provides the true, noise-free SSH.

2. An equally serious objection is that, if the authors' conceptual model is correct, then the estimated squared coherence should be independent of the mapping system. (This is clear from Eqn 6 of the paper.) This point needs to be emphasized: if the assumption of a Case 2 model is correct, then the coherence should only depend on the properties of the input data, and there should be no difference between different mapping schemes (but these differences are much of what the paper is about).

3. Another related objection is that it doesn't seem like the authors are clear about what they mean by resolution. It seems the authors are saying the resolution is defined as the wavenumber at which the noise in the input data equals the signal in the input data. This doesn't make sense for at least a couple of reasons:

   (a) The resolution capability of a mapped field should depend on (i) the noise and signal levels in the input data, (ii) the sampling of the data, and (iii) the manipulations performed on the data during the mapping (e.g., filtering).

   (b) The conceptual model of resolution doesn't seem to have anything to do with the mapping system (item iii directly above) or the sampling (item ii directly above), which are clearly important factors affecting the resolution of the data product. The measure of resolution used here only depends on the SNR of the along-track input data. Admittedly, the along-track measurement noise could be important. But, in fact, important previous work on the resolution of altimetry maps, like Wunsch (1989), and Greenslade/Chelton/Schlax (1997) have neglected measurement noise for the most part, arguing that sampling errors and the filtering inherent in making the maps (and needed to reduce sampling errors) are the dominant factors determining resolution of the mapped data. Another way to say this is that, even if the noise on the input is zero, the resolution will still be limited because of the information lost in sampling and the filtering that is inherent in the mapping system.

   (c) It seems like the filtering properties of the system should be considered in discussing resolution. For example, if a version of the data was low-pass filtered, I think the authors would agree it has lower effective resolution than the input data. This coherence-based criterion for defining resolution does not necessarily have any clear relationship to the filtering properties of the mapping system.

I have prepared three examples to illustrate the behavior of the proposed measure of resolution and to compare it to more commonly used measures. In all three examples, the measurement noise is zero. (It could be nonzero, but it is clearer to set the noise to zero.) The examples also have no sampling errors. The examples thus focus on how the filtering properties of the mapping algorithm affect the proposed measure of resolution. The three examples are: (1) a 1-D case, (2) a 2-D case with a white spectrum, and (3) a 2-D case with a red spectrum. In all three examples, I mapped the data using three different smoothers, a Gauss-Markov smoother (also known as optimal interpolation), a Gaussian weighted average smoother, and a quadratic loess smoother. At least a few of the authors are familiar with these smoothers, and the exact details of the smoothers are not important. The loess and Gaussian smoother parameters were chosen such that they filter the data with a 25-km half-power filter cutoff in one spatial direction (nominally the "along-track" direction). The Gauss-Markov smoother uses a Gaussian autocovariance function and assumes the measurement contains a small amount of white noise, and it has similar but not identical filtering properties (with an autocovariance function temporal decay timescale of 25 days).

1 1-D example

In the first example, the input signal is a random realization of a process having a red spectrum (k^{-2} power law), sampled on a uniform spatial grid. The input data were "mapped" (smoothed) to the same grid as the sampling positions. There is no measurement noise and no sampling error. In this case, the only thing limiting the resolution of the mapped fields is the filtering inherent in the mapping algorithm. One could imagine this as a case where there is a single along-track pass of data, and they have been mapped to a regular spatial grid (along-track).

Figure 1 (top) shows the spectrum of the input data and of the three mapped fields. The mapped fields have less variance at high wavenumbers because of the filtering inherent in the mapping.

The filtering inherent in the mapping is quantified more directly by computing the cross-spectral gain (or relative amplitude of variability coherent between the input data and the mapped data). The gain is equivalent to the magnitude of the filter transfer function of the filters, and is shown in Figure 1 (middle panel). The Gaussian smoother has a filter transfer function that resembles a Gaussian function (as is expected, because the Fourier transform of the spatially truncated Gaussian weighting function is approximately a Gaussian). The loess smoother has, as expected, a steeper filter roll-off (meaning it decreases from one toward zero more abruptly near the half-power point) with a small but noticeable filter sidelobe at about twice the half-power cutoff wavenumber (Schlax and Chelton, 1992, their Figure 1). The Gauss-Markov smoothed estimate has a very steep roll-off, and it has a half-power filter cutoff that is at a slightly higher wavenumber than the loess and Gaussian smoothed fields. All of these same features can be clearly seen in the spectra, as well (upper panel of Figure 1).

Based on conventional understandings of the term "resolution", we would say that the resolution, as defined by the half-power point of the filtering, is highest in the Gauss-Markov mapping and is the same in the loess and Gaussian weighted average mappings. (One might reasonably argue the resolution of the Gaussian mapping is worse because it noticeably attenuates wavelengths longer than the half-power cut-off wavelength.) For example, this perspective is similar to the definition of resolution used by the SWOT project when they state the raw (downlinked) resolution of the in-swath SWOT data will be 1-km resolution and 500-m posting– the onboard processor will filter the data, which fundamentally limits the resolution.

Now, if we turn our attention to the squared coherence (bottom panel of Figure 1), and apply

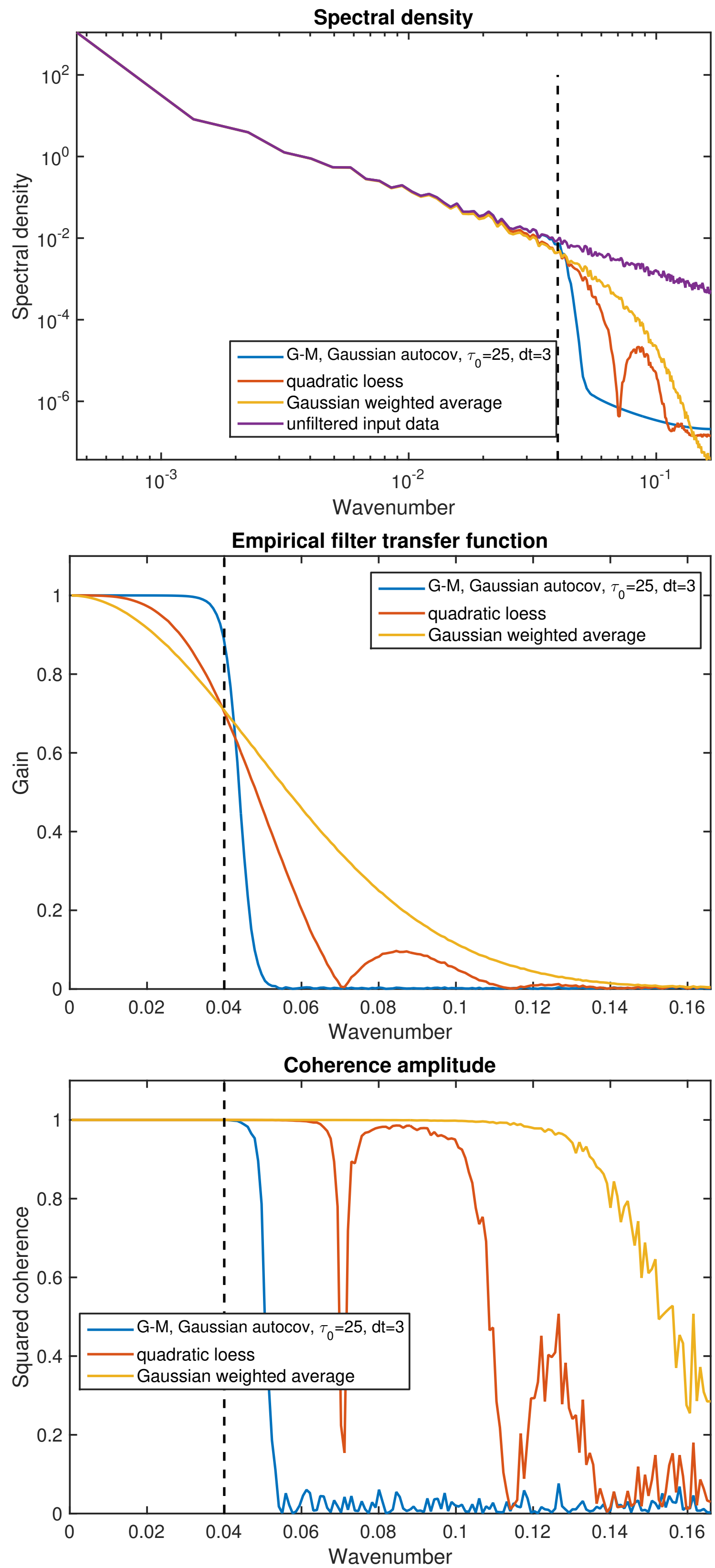

Figure 1: 1D example. Upper panel: Spectra of input signal and mapped fields. Middle panel: squared coherence of the mapped field with the input signal. Lower panel: spectral gain computed between the input signal and the mapped field. The black dashed line marks the theoretical half-power wavenumber of the loess and Gaussian smoothers.

the authors' coherence-based definition of resolution, we would reach the opposite conclusion: the Gaussian smoother has vastly superior resolution, and the squared coherence is almost one, even at wavenumbers three times larger than the half-power point, where the variance is more than 100 times less than the actual (unfiltered) variance. Under the definition of resolution proposed by the authors, the resolution of the mapping would be 6.5km wavelength for the Gaussian mapping, 14.3km wavelength for the loess mapping, and 20km wavelength for the Gauss-Markov mapping. The maximum possible resolution would be 6km wavelengths because the sample spacing in this example is 3km.

I had actually expected the coherence in this example to be one at all wavenumbers for all three mappings. We should expect this because all of the mapping algorithms are linear, meaning that there should be a perfectly linear relationship between the amplitude and phase of the input and output at all wavenumbers, and thus a coherence of one at all wavenumbers. I believe the reason it is not one has to do with numerical errors (such as roundoff errors) that occur at the higher wavenumbers where the amplitude of the filtered signal is very small.

In summary, this first example shows that the coherence-based measure of resolution fails to be useful in a simple, but relevant, example. The example is relevant because the filtering of the mapping procedure is an important aspect of the resolution of the mapped field, and the coherence criterion tells us almost nothing about this filtering.

2 2-D example with a white spectrum

I imagine the authors might wonder, as I did when thinking about the above example, why the coherence-based measure of resolution seems to provide reasonable results when applied to the DUACS system. The next two examples show another problem with the coherence-based measure of resolution that is related to the fact that the filtering in one dimension (e.g., time) has important effects on the coherence in the other dimension (e.g., space).

In this second example, the input signal is a random realization of a process having a white spectrum, sampled on a uniform 2-D grid meant to represent space in one dimension and time in the other. The input data were "mapped" (smoothed) to the same space-time grid as they were sampled on. There is no measurement noise and no sampling error. In this case, the only thing limiting the resolution of the mapped fields is the filtering inherent in the mapping algorithm. One could imagine this case as a case where there are repeated along-track passes of data, and they have been mapped to a regular space-time grid (along-track). Alternatively, instead of thinking of one dimension as time, we could imagine the two dimensions as along- and across-track directions. (Please note that these examples have noisier and more coarsely resolved spectral estimates because I could not afford the computer power/time to use a very large number of samples.)

In this second example, the smoothing in the "along-track" direction is the same as in the 1-D example. In the other dimension (call it time), there is some smoothing (10-unit half-power wavelength for the loess and Gaussian smoothers, and similar for the Gauss-Markov mapping), but the exact value doesn't matter.

In analogy to the approach used in the paper, I took a single "along-track" sample of the input data and of the 2D mapped data (from the central "time" of the mapped domain, to avoid edge effects in the mappings). The spectra of the along-track input data and mapped data are shown in Figure 2 (upper panel). The variance of all three mapped fields is reduced at all wavenumbers relative to the input data, but the variance reduction is greatest at the highest wavenumbers.

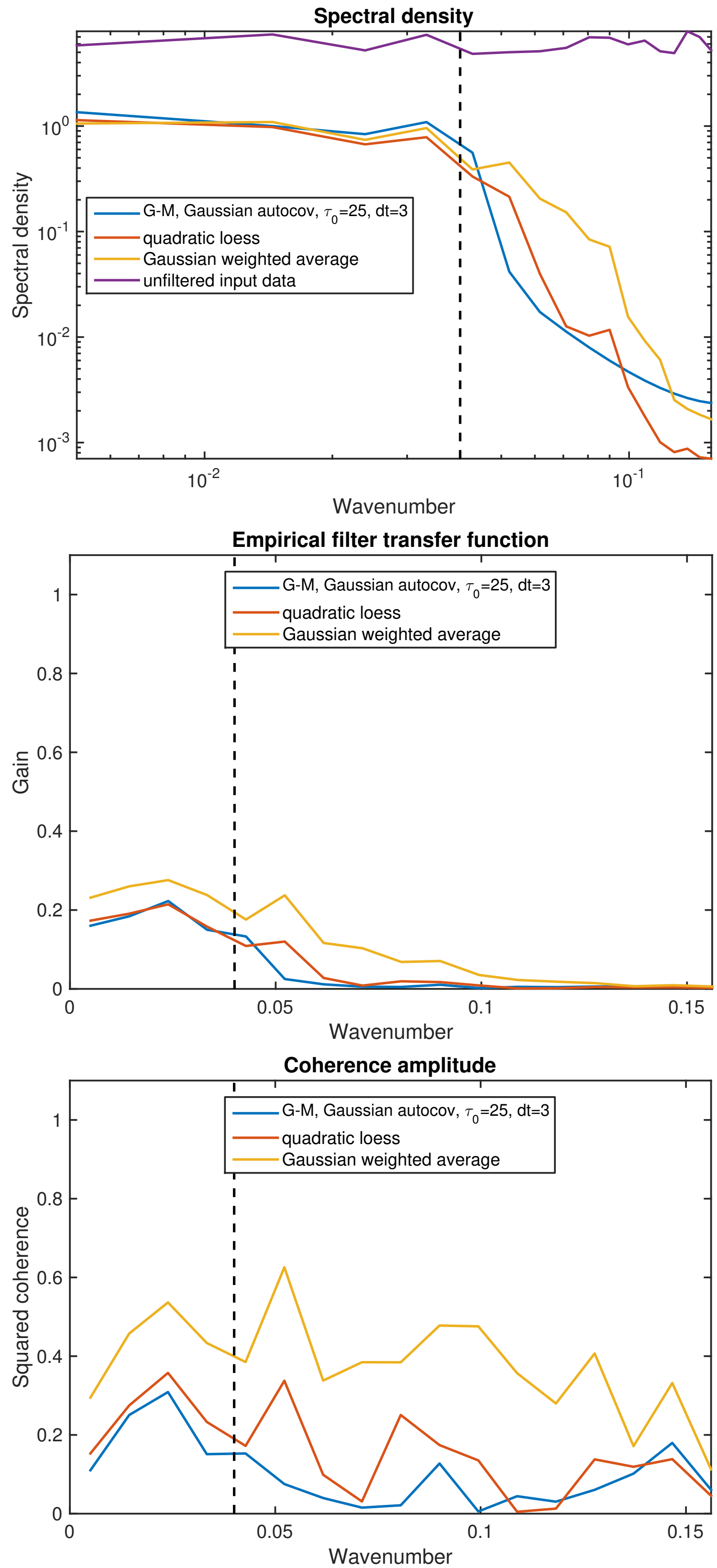

Figure 2: 2D example for a variable that has a white spectrum. Upper panel: Spectra of along-track samples of input signal and mapped fields. Middle panel: squared coherence of the mapped field with the input signal. Lower panel: spectral gain computed between the input signal and the mapped field. The black dashed line marks the theoretical half-power wavenumber of the loess and Gaussian smoothers in the along-track direction.

This is easy to understand– it just reflects the fact that the along-track filtering affects the larger along-track wavenumbers but the temporal filtering affects all along-track wavenumbers equally and accordingly causes a uniform reduction in variance. (This is analogous to how the across-track smoothing of SWOT data should reduce the contributions of noise to the along-track spectrum.)

The estimates of gain between the along-track input data and mapped data tell a similar story (Figure 2, middle panel). The mapped fields are attenuated relative to the input data at all along-track wavenumbers because of the temporal filtering, and they are even more attenuated at the high along-track wavenumbers because of the along-track filtering.

The squared coherence is not especially interesting. Squared coherence in a particular wavenumber band can generally be interpreted as the fraction of the variance at that wavenumber in the input record (the raw along-track data) that can be accounted by multiplying the Fourier transform of the output record (the along-track mapped data) by some complex-valued constant (the value of the transfer function at that wavenumber). So, we can see another aspect of what we saw in the spectra and gain plots– the coherence is reduced at all wavenumbers because the temporal filtering reduced the variance in the mapped fields at all wavenumbers relative to the raw along-track record, and thus there is only limited ability of the along-track mapped data to account for the variance in the raw along-track data. (I struggled to understand this, but I found this simple example helpful: imagine if the temporal averaging were very extreme such that the mapped data is very close to the time-mean SSH anomaly; in that case, we would expect very low along-track coherence with an along-track pass of SSH anomaly that would be dominated by eddies and variability.)

If we did try to apply the coherence-based measure of resolution in this example, we would conclude that the wavelength resolution of the Gauss-Markov and loess mappings is larger than the domain size (1800 km), and that the Gaussian mapping might resolve about 50 km wavelengths.

3 2-D example with a white spectrum

I imagine the authors still might wonder, as I did when thinking about the above example, why the coherence-based measure of resolution seems to provide reasonable results when applied to the DUACS system. The coherence in Figure 1 of the paper, for example, does not look like the coherence in Figure 1 or in Figure 2. Instead, Figure 1 of the paper has high coherence at low wavenumbers and low coherence at high wavenumbers. I think the explanation is that the variabilty in SSH has a red spectrum in space and time, so that the low wavenumbers tend to be associated with energetic low frequencies, and the low frequencies tend to be associated with energetic low wavenumbers. This example, with a signal that has a red spectrum in both wavenumber and frequency, is meant to illustrate how the coherence-based measure of resolution applied to a red signal spectrum combined with the multiple-dimension filtering of the DUACS system can lead to results that seem reasonable, even though the apparently reasonable looking results are basically accidental.

In this third example, the input signal is a random realization of a process having a red spectrum in both wavenumber and frequency (spectrum proportional to $k^{-2}\omega^{-2}$), sampled on a uniform 2-D grid meant to represent space in one dimension and time in the other. All other aspects are identical to the second example.

I again took a single "along-track" sample of the input data and of the 2D mapped data. The

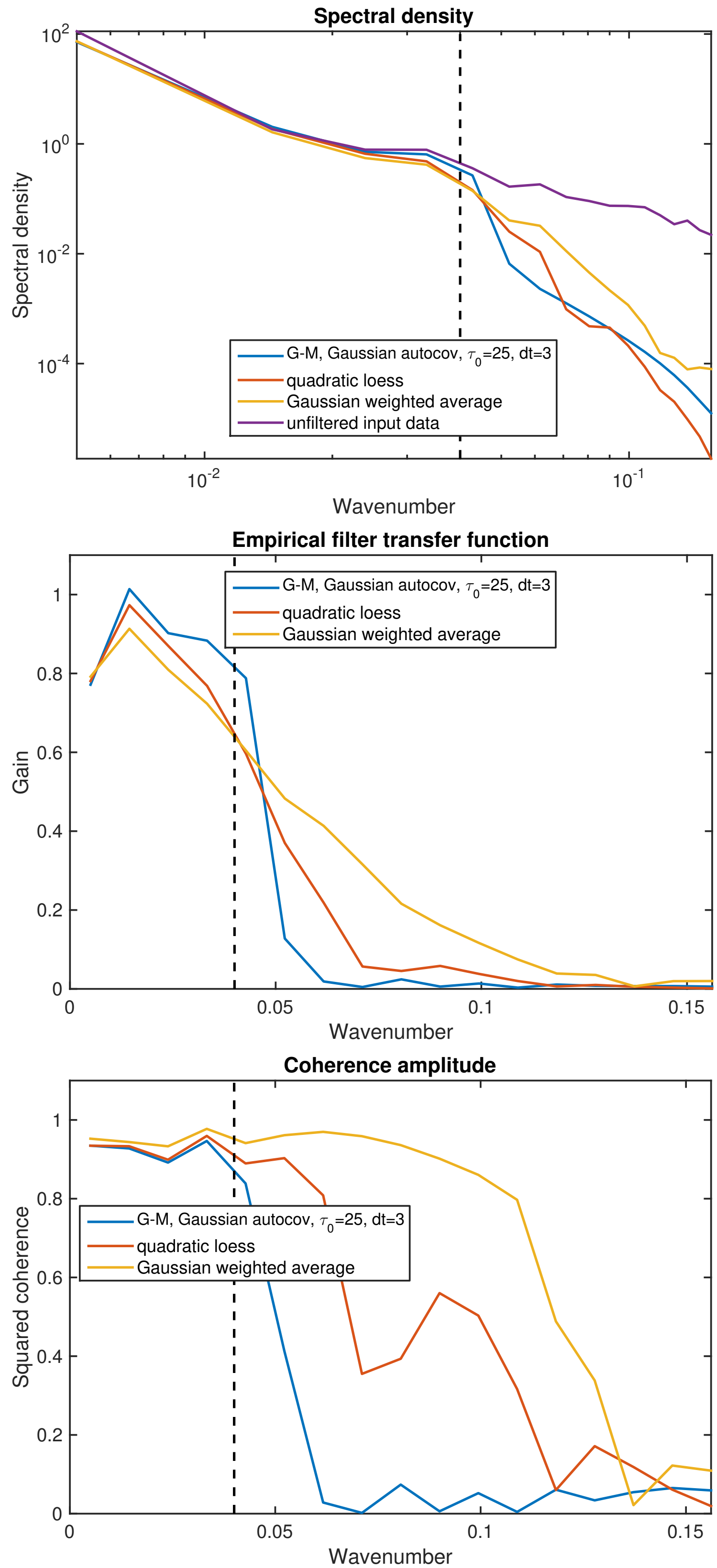

Figure 3: 2D example for a variable that has a spectrum that is red in wavenumber and frequency. Upper panel: Spectra of along-track samples of input signal and mapped fields. Middle panel: squared coherence of the mapped field with the input signal. Lower panel: spectral gain computed between the input signal and the mapped field. The black dashed line marks the theoretical half-power wavenumber of the loess and Gaussian smoothers in the along-track direction.

spectra of the along-track input data and mapped data are shown in Figure 3 (upper panel). Unlike the second example with white spectra, the variance of the three mapped fields is not noticeably reduced at low wavenumbers. This is not difficult to understand– there is relatively little variance at the high frequencies in the temporal dimension, so the temporal low-pass filtering associated with the mapping has relatively little effect on the variance in the along-track direction.

The gain plots in this example (Figure 3, middle panel) look more similar to the 1-D example than to the 2-D example with a white noise signal. The half-power points one would infer from the gain are at similar wavenumbers to the theoretical 25km half-power point of the loess and Gaussian smoothers (29km wavelength for the loess, 27km wavelength for the Gaussian, and 22.7km wavelength for the Gauss-Markov mapping). This good agreement is totally accidental. If the frequency spectrum of the SSH were different, or if the temporal smoothing were different, the gain between the along-track input data and mapped data would change, and the point where it is equal to $\sqrt{0.5}$ (i.e., the half-power point) would be different.

The squared coherence in this example looks qualitatively similar to Figure 1 of the paper, with high coherence at low along-track wavenumbers and low coherence at high wavenumbers. The coherence-based definition of resolution would yield along-track wavenumbers of 8.5km wavelength for the Gaussian mapping, 14.7km wavelength for the loess mapping, and 19.9km wavelength for the Gauss-Markov mapping.

4 Conclusion

The along-track filtering properties of the mapping schemes should be the same in all three examples. (For example, we can analytically derive the filtering for the Gaussian weighted average.) The half-power point of the filtering would not be a good specification of the resolution, in general, because the resolution will also depend on the sampling and on the SNR of the input data. However, in these examples, which do not have any noise or sampling errors, the effective resolution should be determined only by the filtering. The filter half-power points of the along-track filtering were at a wavelength of about 25km for all three mapping schemes.

I suppose all measures of resolution have their drawbacks (e.g., the spectral ratio approach discussed in the paper is subject to some of the same issues, such as sensitivity of inferred along-track resolution to temporal filtering), but I do not see any theoretical basis for the coherence-based measure of resolution. In the cases we considered, the coherence-based measure yielded along-track wavelength resolutions ranging from 6.5km (close to the Nyquist wavelength) to >1800km (the domain size) for cases in which the actual resolution was about 25km, and I am confident that almost any result could be obtained by varying the signal spectrum and/or the temporal filtering. In a case with no noise and no sampling errors, the along-track wavenumber resolution should not depend on the signal spectrum or the temporal filtering. I cannot think of any relevant case in which it would make sense to define the resolution as the wavenumber where the squared coherence between the mapped product and a one-dimensional sample of the input data is equal to 0.5.

I think that the conclusions of the paper are not quantitatively useful and that it would be counterproductive to publish this. I feel really bad having to say that, because it is clear the authors worked hard to make a high-quality paper. Except for the one methodological flaw, it is an excellent paper in all other respects.

I have focused on the fact that the coherence-based measure of resolution does not take into account the filtering by the mapping system. The coherence-based measure of resolution also does not adequately take account of the sampling, and this might also be a serious issue.