Ludwig-Maximilians-Universität München
Faculty of Physics

Master-Thesis

Field dynamics inference via spectral density estimation

supervised by:
PD Dr. Torsten Enßlin
Max Planck Institute for Astrophysics

submitted by:
Philipp Frank
10753005
13.03.2018

Ludwig-Maximilians-Universität München
Fakultät für Physik

Masterarbeit

Felddynamik-Inferenz via spektrale Dichte-Schätzung

betreut von:
PD Dr. Torsten Enßlin
Max-Planck-Institut für Astrophysik

vorgelegt von:
Philipp Frank
10753005
13.03.2018

Table of Contents

1 Introduction

[1] Stochastic differential equations (SDEs) play an important role in a variety of scientific fields [2] and industrial designs [3]. SDEs are a flexible tool to model dynamical processes and traditionally serve as a useful prior for the inference of dynamical fields. In recent decades the inverse problem for SDEs gained considerable attention as well. The goal of this inverse problem is to infer the dynamical properties of the system from observations of an evolving field. In order to tackle this problem a variety of methods have been proposed in the past [4, 5]. These range from parametric methods which aim to parametrize the process either in the temporal [6] or the frequency domain [7] to non-parametric and Bayesian methods [8–10].

In a number of physical disciplines (e.g., astrophysics or plasma-physics) the dynamical quantities of interest may evolve not only in time but are also extended in space. To infer the underlying spatio-temporal dynamical process, a variety of methods have been proposed. These include kernel-based methods [11], approaches using spatio-temporal Kriging [12] as well as Bayesian non-parametric approaches for multi-dimensional spatial evolution [13]. Some methods aim to tackle this problem assuming separability of the spatial and the temporal evolution which helps to simplify the inference problem. In many physical applications, however, separability cannot be assumed due to the entanglement of structures in space and time.

For autonomous processes, the dynamics is fully determined in terms of the spectral density of the dynamical field. In this work we will use this property to model linear, non-separable, SDEs. Furthermore, the proposed method is able to reconstruct the spectral density also for noisy and masked observations of the dynamical field using Bayesian inference.

Physical fields usually can be defined in a corresponding configuration space and linear differential equations can be described as linear operators acting on field vectors in this space. We model the problem with the help of probabilities over the elements of this configuration space directly to reflect the mentioned properties of the field. To this end, we rely on the language of information field theory (IFT) [14, 15] which extends information theory to fields. The restriction of this work to spatially and temporal autonomous differential operators enables us to describe those in terms of fields over the joint Fourier spaces and to derive posterior distributions for the corresponding fields directly. Due to the fact that the resulting equations are often not analytically traceable, the computations ultimately are performed on a finite grid on a computer. However, as IFT is independent of the chosen discretization, one can always choose a representation which is convenient for the problem at hand. In this work, the discretization is achieved using the software package NIFTy 3 (Numerical Information Field Theory) [16] which allows

[1] Major parts of sections 1 - 4 and 6 were already published in [1]

building algorithms using a field theoretical language, while the framework takes care of the underlying discretization. In order to introduce the notation and field theoretical language used throughout this work, we continue with a brief introduction to IFT.

1.a Introduction to IFT and notation

IFT is a statistical field theory which allows doing probabilistic calculations for fields (in the physical sense), defined over continuous space (or space-time). It is already used in various inference tasks such as astrophysical imaging [17], component separation [18], numerical simulations of dynamical systems [19] and others. IFT has also previously been applied in the context of dynamical system inference [20]. The theory enables the user to work directly in the field configuration space, which for many applications involving fields is the natural space to define problems.

In order to define probabilities for fields we introduce a scalar product for fields ($\phi(x)$, $\psi(x)$ with $x \in \Omega \subset \mathbb{R}^n$) as

$$\phi^\dagger \psi = \int_\Omega \phi^*(x)\psi(x)\, dx \ , \tag{1}$$

where * denotes complex conjugation. Note that throughout this work we restrict ourselves to scalar fields, for simplicity, although a generalization to vector fields is possible. As an example, a Gaussian distribution for a field can be written as

$$\begin{aligned}\mathcal{P}(\phi) &= \mathcal{G}(\phi - m, \Phi) \\ &= \frac{1}{\sqrt{|2\pi\Phi|}} \exp\left\{-\frac{1}{2}(\phi - m)^\dagger \Phi^{-1}(\phi - m)\right\} \ ,\end{aligned} \tag{2}$$

where m denotes the mean and Φ is a linear, self-adjoint and strictly positive operator which maps from the field configuration space to itself. In other words, it is the continuous version of a covariance. Note that Φ is sometimes referred to as a covariance matrix, although it is actually a continuous operator. $|\Phi|$ denotes the functional determinant of the operator Φ. The exponential factor in Eq. 2 reads

$$\phi^\dagger \Phi^{-1} \phi = \phi^*_x \Phi^{-1}_{xy} \phi_y = \int dx \int dy \ \phi^*(x)\ \Phi^{-1}(x,y)\ \phi(y) \ , \tag{3}$$

where we also introduced the continuous version of the Einstein sum convention.

Although it is natural to formulate many physical problems in a field theoretical language, the data that we observe is always finite. Therefore, we need to translate between the field configuration space (called signal space) and the so called data space. To illustrate this procedure consider the following measurement scenario

$$d_u = R_u[\phi] + n_u \ , \tag{4}$$

where R denotes a projection operator which projects from the infinite dimensional configuration space of ϕ to a finite set of M measurement points d_u (with $u \in \{1, .., M\}$) and n is the measurement noise. Using this data model and Bayes theorem, the posterior of ϕ given d reads

$$\mathcal{P}(\phi|d) \propto \int \mathcal{D}n \; \mathcal{P}(d|\phi, n)\mathcal{P}(\phi)\mathcal{P}(n) \; . \tag{5}$$

If we assume n and ϕ to be independently Gaussian distributed with zero mean and covariances N and Φ, respectively, and also assume R to be a linear operator the posterior is also a Gaussian. I.e.

$$\mathcal{P}(\phi|d) = \mathcal{G}(\phi - m, D) \; , \tag{6}$$

with mean

$$m = Dj = (\Phi^{-1} + R^\dagger N^{-1} R)^{-1} R^\dagger N^{-1} d \; , \tag{7}$$

and covariance $D = (\Phi^{-1} + R^\dagger N^{-1} R)^{-1}$. These equations resemble the famous Wiener filter equations [21] which is the optimal linear filter, applied to a field theoretical setting.

In a non-linear setting, however, the exact form of the posterior or corresponding expectation values are often not analytically traceable. Therefore, in this work we rely on maximum a posteriori (MAP) estimates which we obtain by minimizing the so called information Hamiltonian defined as

$$\mathcal{H}(x) = -\log\left(\mathcal{P}(x)\right) \; , \tag{8}$$

where $\log(x)$ is the natural logarithm. In addition, the second derivatives of the Hamiltonian give rise to the Laplace approximation of the uncertainty maps, as we will discuss in section 3.e.

To motivate the application of IFT to the inference of dynamical systems we note that linear differential equations can be rewritten as linear (differential) operators acting on a field of interest. As we will point out in the next section, these operators serve as a building block for the covariance function in the context of SDEs. Specifically, we draw the connection to the spectral density of the field which is defined as the diagonal of the covariance operator in harmonic space.

1.b Structure of this work

The rest of this work structures as follows: In section 2 we briefly outline how SDEs are connected to the spectral density of a random process. Consequently, in section 3 we describe the key properties of our inference method of spectral densities from field realizations as well as noisy measurements thereof. In section 4 we apply our method to different mock data examples including one- and two-dimensional examples. In section 5 we use our method to reconstruct fluctuations in the electron density of the ionosphere

using data from the LOw Frequency ARray (LOFAR [22]) radio interferometers and show preliminary results. Finally, in section 6, we conclude the thesis with a short summary as well as a small outlook to possible applications and further projects.

2 From a SDE to the spectral density

In this section we outline how the properties of a linear SDE are encoded in the spectral density of a spatio-temporal dynamical process. Furthermore, we introduce the key assumptions that are necessary to ensure that all relevant information is encoded in the density.

A suitable starting point is a linear SDE of the form:

$$(\mathcal{L}\phi)(\mathbf{x}, t) = \xi(\mathbf{x}, t) \ , \tag{9}$$

where $\mathcal{L}$ is a linear (differential) operator acting on a field ϕ, and ξ is a random process. $\mathbf{x} \in \mathbb{R}^d$ and $t \in \mathbb{R}$ denote the spatial and temporal coordinates, respectively. Note that the distinction between space and time is for convenience only, i.e. all formulas treat space and time on the same footage which means that the analysis is also valid for a general multi-dimensional process.

Assuming ξ to be Gaussian distributed with a covariance matrix Θ and using Eq. 9 yields the probability distribution for ϕ:

$$\mathcal{P}(\phi|\mathcal{L}, \Theta) = \int \mathcal{D}\xi \ \mathcal{P}(\phi|\mathcal{L}, \xi) \ \mathcal{G}(\xi, \Theta) = \mathcal{G}(\phi, \Phi) \ , \tag{10}$$

where

$$\Phi = (\mathcal{L}^\dagger \Theta^{-1} \mathcal{L})^{-1} \ . \tag{11}$$

As we can see, all relevant information concerning the statistical properties as well as the dynamic evolution is encoded in the correlation matrix Φ if the underlying process is linear and invertible.

Assuming $\mathcal{L}$ to be local and homogeneous in both, space and time, implies that the operator can be written as

$$\mathcal{L}_{xx'} = \delta^{(d+1)}(x - x') \ g(\partial_t, \partial_{\mathbf{x}}) \ , \tag{12}$$

where we introduced the space-time vector $x = (t, \mathbf{x})$ and the differential operator encoding function g. Fourier transforming Eq. 12 yields:

$$\mathcal{L}_{kk'} = (2\pi)^{d+1} \delta^{(d+1)}(k - k') \ f(k) \ , \tag{13}$$

where $k = (\omega, \mathbf{k})$ denotes the coordinates in harmonic space and $f(k) = g(i\ \omega, i\ \mathbf{k})$ is a

complex scalar field, with i being the imaginary unit. Note that if the differential equation is real, then $f^*(\omega, \mathbf{k}) = f(-\omega, -\mathbf{k})$ and therefore $\mathcal{L}$ is Hermitian.

Assuming further that also Θ is diagonal in harmonic space with the spectral density $P_\xi(k)$, Eq. 11 can be rewritten as

$$\begin{aligned}\Phi_{kk'} &= (2\pi)^{d+1}\delta^{(d+1)}(k-k')\frac{P_\xi(k)}{|f(k)|^2} \\ &=: (2\pi)^{d+1}\delta^{(d+1)}(k-k')P_\phi(k)\ ,\end{aligned} \tag{14}$$

where we defined $P_\phi(k)$, the spectral density of ϕ.

As we can see, P_ϕ encodes the properties of a SDE up to the complex phase of f. Therefore, we seek to find a way to infer it from observations of the field ϕ. In order to derive the posterior distribution of the spectrum, in the following, we propose a way to model the key features of P_ϕ.

3 Spectral density inference

In order to model P_ϕ we notice that if f and P_ξ are continuous and smooth functions of their arguments, then P_ϕ is a rational and positive function. We therefore model the spectral density as

$$P_\phi(k) = \exp\left[\tau(k) + \tan\left(\delta(k)\right)\right]\ . \tag{15}$$

The idea behind this definition is that we want to reduce P_ϕ to its two key properties: either P_ϕ is a smooth, positive function of k, which we model by $\exp(\tau)$, or it diverges as

$$|f(k)|^2 \to 0\ , \tag{16}$$

which is modeled by $\exp(\tan(\delta))$. We therefore define suitable prior distributions for τ and δ which aim to support these features.

3.a τ-Prior

In order to constrain τ to be a smooth function of its arguments, we impose a smoothness prior on τ (see e.g., [23]). To get an idea about smoothness and the corresponding prior consider the following 1D example: Suppose P_ϕ models a power-law, i.e.

$$P_\phi(y) = y^\alpha\ , \tag{17}$$

with $y, \alpha \in \mathbb{R}$ and assume for the moment $\delta = 0$. Then

$$\tau = \log\left(P_\phi(y)\right) = \alpha\log\left(y\right)\ , \tag{18}$$

is linear in $\log(y)$ which implies that the second logarithmic derivative vanishes. We therefore built our prior such that it minimizes the curvature of τ on a logarithmic scale. This reads

$$\mathcal{P}(\tau) = \mathcal{G}(\tau, T_\sigma) \ , \tag{19}$$

with T_σ such that

$$\tau^\dagger T_\sigma^{-1} \tau = \frac{1}{\sigma^2} \int d\left(\log\left(y\right)\right) \left| \frac{\partial^2 \tau(y)}{\partial \log\left(y\right)^2} \right|^2 \ , \ \sigma \in \mathbb{R} \ , \tag{20}$$

where σ is an overall hyper-parameter controlling the degree of smoothness one wants to impose on τ.

In higher dimensions, this constraint has to be imposed for all quadratic, logarithmic variations of the field simultaneously. A derivation of the exact form as well as a short discussion can be found in appendix A. Furthermore, in some applications it is necessary to impose smoothness also for negative y. To do so, we extend this prior using the complex logarithm, which is also defined on a negative scale. Details of this approach as well as the treatment of the special point $y = 0$, are discussed in appendix B.

3.b δ-Prior

The prior distribution for δ is constructed in a way such that δ allows for a transition of the spectral density from smooth to divergent regions. We therefore also impose a smoothness prior on δ which implies that for small δ, where

$$\tan(\delta) \approx \delta \ , \tag{21}$$

the spectrum remains smooth. However, as δ approaches $\pm\pi/2$, small changes in δ result in large, abrupt changes of the spectrum.

The full prior reads

$$\mathcal{P}(\delta) \propto \mathcal{G}(\delta, T_\mu) \ \mathcal{G}(\delta, \nu^2 \mathbb{1}) \ , \ \delta \in [-b, b] \ , \ \mu, \nu \in \mathbb{R} \ , \tag{22}$$

where we also included a term to the prior that punishes larger values of δ. This ensures that δ remains zero in regions where the data does not support a divergence. Note that we restrict the support of δ to $b = (\pi/2 - \epsilon)$, where ϵ serves as a "high-energy" cutoff in order to avoid infinities during reconstruction. Since the length-scale of δ is $\pi/2$ we note that $\nu \approx \pi/2$ is a reasonable choice.

3.c Perfect data Posterior

Using the priors and the likelihood, defined in Eq. 10, we can immediately write down the posterior distribution

$$\mathcal{P}(\tau, \delta|\phi) \propto \mathcal{G}(\phi, \Phi)\ \mathcal{P}(\tau)\ \mathcal{P}(\delta) \tag{23}$$

and the corresponding information Hamiltonian

$$\begin{aligned}\mathcal{H}(\tau, \delta|\phi) &= -\log\left(\mathcal{P}(\tau, \delta|\phi)\right) \\ &= \frac{1}{2}\left[\phi^\dagger \Phi^{-1}\phi + \log\left(|\Phi|\right) + \tau^\dagger T_\sigma^{-1}\tau + \delta^\dagger\left(T_\mu^{-1} + \nu^{-2}\right)\delta\right] + H_0\ ,\end{aligned} \tag{24}$$

where H_0 is a constant that is independent of τ and δ. Minimizing this Hamiltonian with respect to τ and δ leads to their maximum a posteriori estimates, given perfect data on the realization of ϕ.

3.d Noisy data Posterior

In reality, we are usually only able to retrieve noisy measurement data, d, of ϕ. We therefore seek to find a way to infer the spectral density from noisy measurements rather than from ϕ itself. Using the notation introduced in section 1.a and the data model (Eq. 4), the joint distribution reads

$$\mathcal{P}(d, \phi, \tau, \delta) = \mathcal{G}(d - R\phi, N)\mathcal{G}(\phi, \Phi)\mathcal{P}(\tau)\mathcal{P}(\delta)\ . \tag{25}$$

Depending on the measurement process, in particular the form of R, the optimal way to proceed may differ significantly. As this is a general issue concerning Bayesian inference and well discussed in literature, we want to focus the discussion on our definition of the spectral density rather than all possible ways of inference.

However, we note that there exist in principle two different approaches to reconstruction in this case. One way is to minimize the information Hamiltonian corresponding to Eq. 25, with respect to all quantities of interest (ϕ, τ, δ), to obtain a maximum a posterior solution. In cases of high quality data, this is a good way to proceed.

However, in cases of high measurement uncertainty, the high frequency modes of ϕ are suppressed in the reconstruction as they are indistinguishable from the noise. Due to the fact that the natural domain of (τ, δ) is the harmonic domain, the lack of high frequency modes restricts a reliable reconstruction of τ and δ to low frequencies.

A possible way to resolve this issue is to marginalize out ϕ in Eq. 25. For a linear R marginalization is obtained analytically and the marginal information Hamiltonian is

given by

$$\mathcal{H}(\tau,\delta|d) = \frac{1}{2}\left(\log\left(\frac{|\Phi|}{|D|}\right) - j^\dagger D j + \tau^\dagger T_\sigma^{-1}\tau + \delta^\dagger T_\mu^{-1}\delta\right) + \frac{1}{2\nu^2}\delta^\dagger\delta + H_0 \ , \tag{26}$$

with

$$D = \left(R^\dagger N^{-1} R + \Phi^{-1}\right)^{-1} , \ j = R^\dagger N^{-1} d \ , \tag{27}$$

analogous to Eq. 7. Minimizing Eq. 26 leads to the posterior estimates which we will call $\bar{\tau}$ and $\bar{\delta}$ in the following.

In the spirit of the empirical Bayes approach [24], where we treat these estimates as the true values irrespective of their corresponding uncertainties, the approximate posterior of ϕ reads

$$\mathcal{P}(\phi|d,\bar{\tau},\bar{\delta}) = \mathcal{G}(\phi - \bar{D}j, \bar{D}) \ , \tag{28}$$

where $\bar{D}$ denotes the information propagator D as given by Eq. 27, evaluated at $\bar{\tau}$, $\bar{\delta}$.

3.e Hamiltonian gradients and curvature

In order to obtain maximum a posterior solutions as well as uncertainty estimates from the information Hamiltonians presented in the previous sections, we need to evaluate the corresponding gradients and curvatures.

For the perfect data Hamiltonian defined in Eq. 24 we find

$$\frac{\partial\mathcal{H}(\tau,\delta)}{\partial\tau_k} = \frac{1}{2}(1 - \phi_k^*\phi_k e^{-\tau_k - \tan(\delta_k)}) + (T_\sigma^{-1}\tau)_k, \tag{29}$$

and

$$\frac{\partial\mathcal{H}(\tau,\delta)}{\partial\delta_k} = \frac{1 - \phi_k^*\phi_k e^{-\tau_k-\tan(\delta_k)}}{2\cos(\delta_k)^2} + \left[\left(T_\tau^{-1} + \frac{1}{\nu^2}\right)\delta\right]_k \ . \tag{30}$$

In order go get an estimate of the spectral uncertainty we have to consider the transformed posterior distribution of τ and $\tan(\delta)$ as they enter the logarithmic spectrum. Specifically,

$$\mathcal{P}(\tau,\tan(\delta)|\phi) = \mathcal{P}(\tau,\delta|\phi)\left|\frac{\delta\tan(\delta)}{\delta\delta}\right|^{-1} \ , \tag{31}$$

where $|\bullet|$ denotes the functional determinant. The corresponding information Hamiltonian reads

$$\mathcal{H}(\tau,\tan(\delta)|\phi) = \mathcal{H}(\tau,\delta|\phi) - \mathrm{Tr}\left(\log\left(\cos(\delta)^2\right)\right) \ . \tag{32}$$

The second derivatives of this Hamiltonian can be used to get a Gaussian approximation of the posterior from which we retrieve an uncertainty estimate for the log-spectrum. The

derivatives read

$$\frac{\partial^2 \mathcal{H}(\tau, \tan(\delta))}{\partial \tau_k \; \partial \tau_q} = \frac{1}{2} \phi_k^* \phi_k e^{-\tau_k - \tan(\delta_k)} \delta_{kq} + (T_\sigma^{-1})_{kq} \; , \tag{33}$$

and

$$\begin{aligned} \frac{\partial^2 \mathcal{H}(\tau, \tan(\delta))}{\partial \tan(\delta_k) \; \partial \tan(\delta_q)} = & \frac{1}{2} \phi_k^* \phi_k e^{-\tau_k - \tan(\delta_k)} \delta_{kq} + \cos^2(\delta_k) \cos^2(\delta_q) \left(T_\tau^{-1} + D_\eta^{-1} + \frac{1}{\nu^2} \right)_{kq} \\ & - 2 \left(\cos^3(\delta) \sin(\delta) \left(T_\tau^{-1} + \frac{1}{\nu^2} \right) \delta - 1 \right)_k \delta_{kq} \; . \end{aligned} \tag{34}$$

The square-root of the diagonal of the inverse operators (which we call $\sqrt{\hat{O}}$) can then be regarded as the one-sigma uncertainty estimate of the corresponding quantity. Further details are described in [23]. For the noisy data posterior (Eq. 26) the derivations are completely analogous.

Since now all ingredients that are necessary for inference are available, consistency tests as well as mock data applications are presented in the next section.

4 Mock data application

For the first consistency check we restrict the analysis to one dimension, the time axis. Consider a differential equation of the form

$$(\alpha \; \partial_t^2 + \beta \; \partial_t + m^2) \; \phi = \xi \; , \tag{35}$$

with $(\alpha, \beta, m^2) = (0.0003, 0.001, 0.5)$. This is the stochastic version of a damped harmonic oscillator. If we assume ξ to be a white noise process with covariance $\Theta = \mathbb{1}$, the spectral density of ϕ becomes

$$P_\phi(\omega) = \frac{1}{(\gamma - \alpha \; \omega^2)^2 + (\beta \; \omega)^2} \; . \tag{36}$$

A signal ϕ (displayed in figure 1 (b)) can then be generated by drawing one sample from the probability distribution corresponding to P_ϕ. Assuming that one is given ϕ, Eq. 24 can be used directly to infer the spectrum of the underlying stochastic process by maximizing the corresponding Hamiltonian. In this application the hyper-prior values are set to $(\sigma, \mu, \nu) = (2.0, 2.0, 0.5\pi)$. The results of the reconstruction are shown in figure 2, where we depict P_ϕ as well as the reconstruction on a log-log-scale. One can see that both fields behave as expected, i.e. τ models the smooth background of the spectrum while δ reconstructs the divergence and is zero everywhere else. In addition, one sees that the assumption of linearity is true up to the divergent part, as expected.

Now we assume that instead of ϕ we are only given noisy and incomplete measurements d of ϕ, as shown in figure 1. We generate mock data assuming Gaussian noise with variance

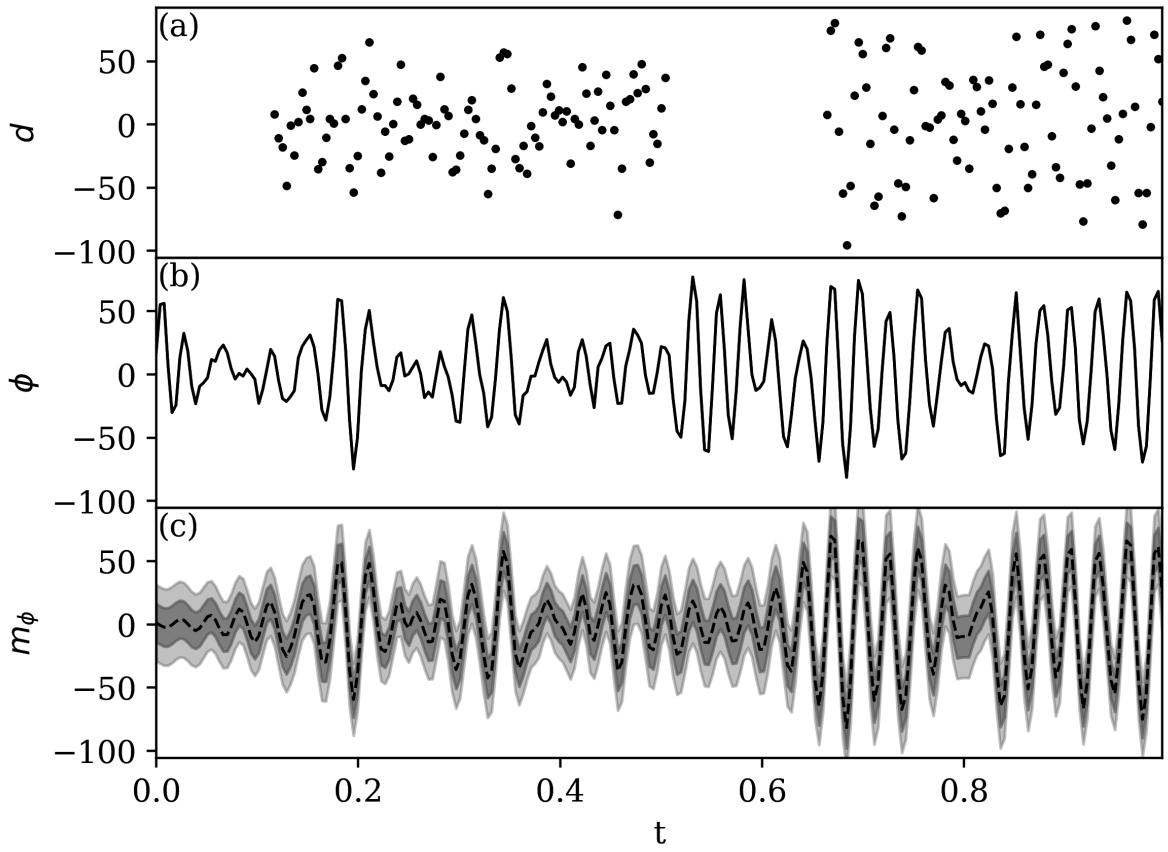

Figure 1: Data d (a), signal ϕ (b) and reconstruction m_ϕ (c) using the MAP estimate of the spectrum. The darker (lighter) gray area denotes the one (two) sigma uncertainty of the reconstruction.

$\sigma_n = 16$. To mimic a more realistic measurement scenario we built the response operator R such that it only measures the signal for certain time intervals. This means that R masks the signal in certain regions and the resulting data, displayed in figure 1 (a), is a noisy and incomplete version of the original signal.

As described in section 3.d, we first reconstruct the spectrum by minimizing the marginal Hamiltonian (Eq. 26) with respect to τ and δ to get maximum a posterior solutions. Thereby we used the same values for the hyper-priors as in the perfect data case. The results are displayed in figure 3. Using the results from Eq. 28 we also obtain a reconstruction of the original signal ϕ as well as corresponding uncertainties (figure 1 (c)). We see that even in regions of no data, we partially infer the correct signal since we were able to obtain a good reconstruction of the spectrum in the first place. The quality of these inter- and extrapolations of the mean reconstruction strongly depends on the correlation length of the corresponding dynamical process. This means that if the process is dominated by random excitations rather than deterministic evolution, interpolation

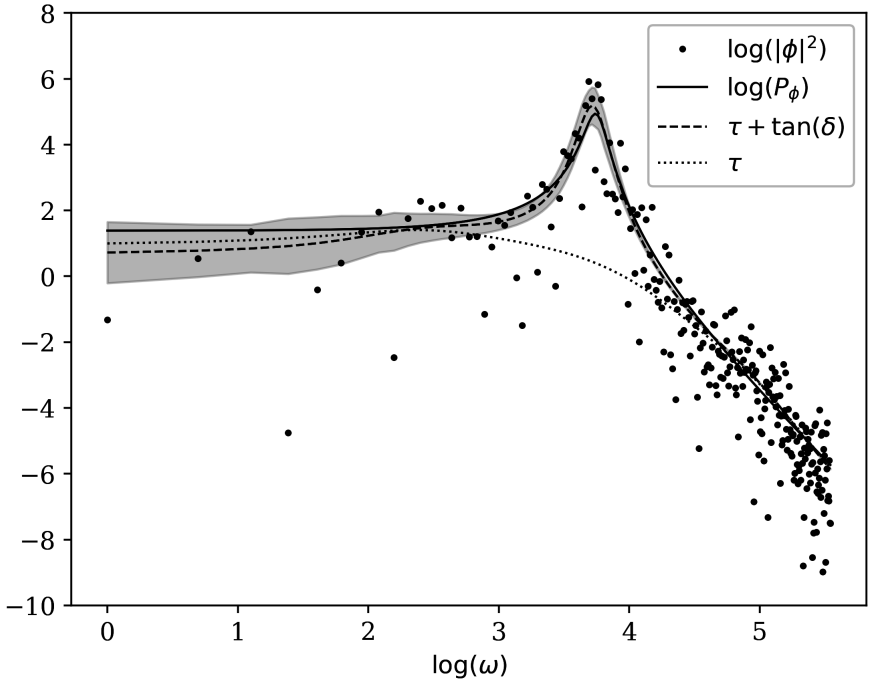

Figure 2: Reconstruction of P_ϕ, given ϕ (figure 1, middle panel), on a natural log-log-scale. The solid line is the theoretical spectrum corresponding to Eq. 36 and the black dots display the spectrum of the actual sample ϕ. The dashed line is the maximum a posterior estimate obtained by maximizing Eq. 24 and the dotted line is the MAP estimate of τ alone. The gray area denotes the one sigma uncertainties of the reconstruction as described in section 3.e.

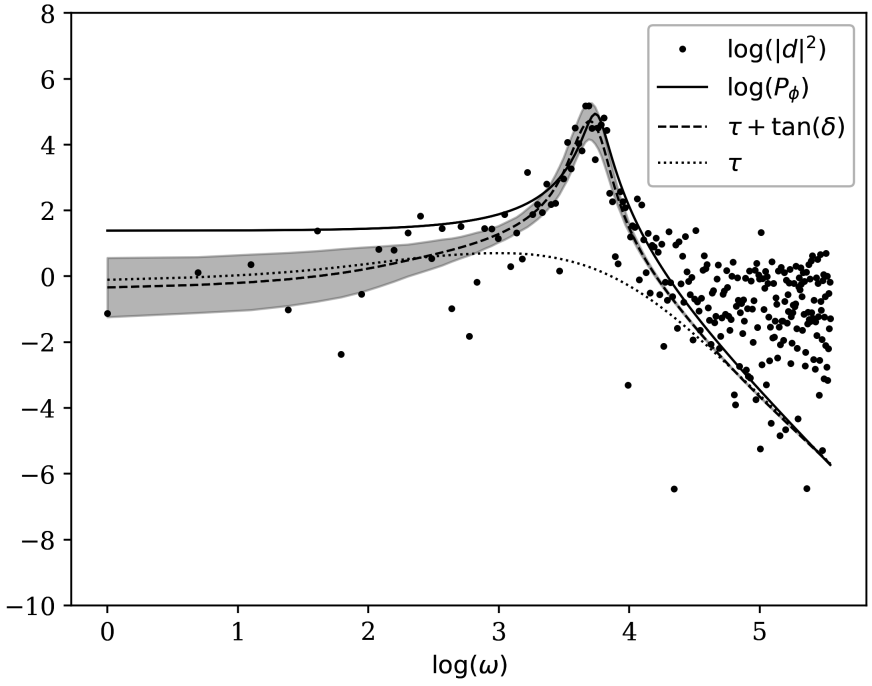

Figure 3: Reconstruction of P_ϕ given noisy data d (figure 1, top panel). We note that due to the Gaussian approximation the uncertainty estimates are significantly underestimated, in regions of low power.

over length scales much larger than the correlation length is in principle not possible. However, due to the fact that we infer the full statistics of the process, even in these cases we are able to state the probability of each possible interpolation in terms of the posterior distribution. Due to the empirical Bayes approach spectral uncertainties are ignored for the reconstruction of ϕ. Therefore, posterior uncertainties of ϕ are underestimated, particularly concerning the overall power of the oscillations. This becomes obvious in the regions of no data.

4.a Spatio-Temporal Evolution

In the next example we extend the analysis to two dimensions, a spatial and a temporal one. The analysis follows the same spirit as described in the previous section. Using a stochastic process of the form

$$(\alpha\partial_t^2 - \beta\partial_x^2 - \gamma\partial_x - \rho\partial_t + m^2)\phi = \xi \ , \tag{37}$$

we first generate mock data d (figure 4 (c)) from a signal ϕ (figure 4 (a)) using a noise variance of $\sigma_n = 7$. ξ is again a white noise process and

$$(\alpha, \beta, \gamma, \rho, m^2) = (0.00007, 0.0002, 0.0014, 0.0012, 0.1) \ . \tag{38}$$

The MAP estimate of the spectrum as well as the spectrum itself is displayed in figure 5. In this case, the hyper-priors are set to

$$(\sigma, \mu, \nu) = (2.5, 2.5, 0.5\pi) \ . \tag{39}$$

We see that in this setting we are able to recover the dominant features of the spectrum, while features of lower power are not recoverable due to noise. This becomes even clearer when we look at figure 7. Here we present slices through the spectrum for different frequency values of k and ω. We see that all features with significant power above the noise level are reconstructed well, while features which are indistinguishable of the noise get suppressed in the reconstruction.

Using the MAP estimate for the spectrum we also reconstruct ϕ itself, as displayed in figure 4 (b). We see that even in regions of no data, we reconstruct the dominant oscillations of the system, while small scale structures cannot be recovered. Again, in figure 6, we present slices of the data, the signal and the reconstruction for different time-steps and at different locations. Subplot (c) shows the spatial structure at a very late time-step, namely in a region where no measurement was made at all. This means that the reconstruction is based entirely on the spectral reconstruction, which serves as a prior, and the constraints which come from data at previous time-steps. Nevertheless,

a reasonable estimate of the field configuration is still possible for a certain period after the last measurements.

To demonstrate the advantage of this non-parametric approach we apply the analysis to a highly structured setting, namely with a spectrum of the form

$$P_\phi(k,\omega) = \frac{2}{\left(m^2 - \sin\left(\alpha k^2 - \beta\omega^2\right)\right)^2 + (\gamma k + \rho\omega)^2} \ , \tag{40}$$

with $(m^2, \alpha, \beta, \gamma, \rho) = (1.1, 0.0025, 0.0011, 0.002, 0.004)$. Although this spectrum is completely artificial, we note that similar periodic and highly-structured spectra also exist in reality. Such are observed, for example, in helioseismology [25]. We use the same setup as described in the previous example but reduce the noise variance to $\sigma_n = 1$ in order to capture more structure of the spectrum. In addition, the hyper-prior parameters were set to $(\sigma, \mu, \nu) = (4.0, 4.0, 0.5\pi)$. The results are shown in figure 8 and figure 9.

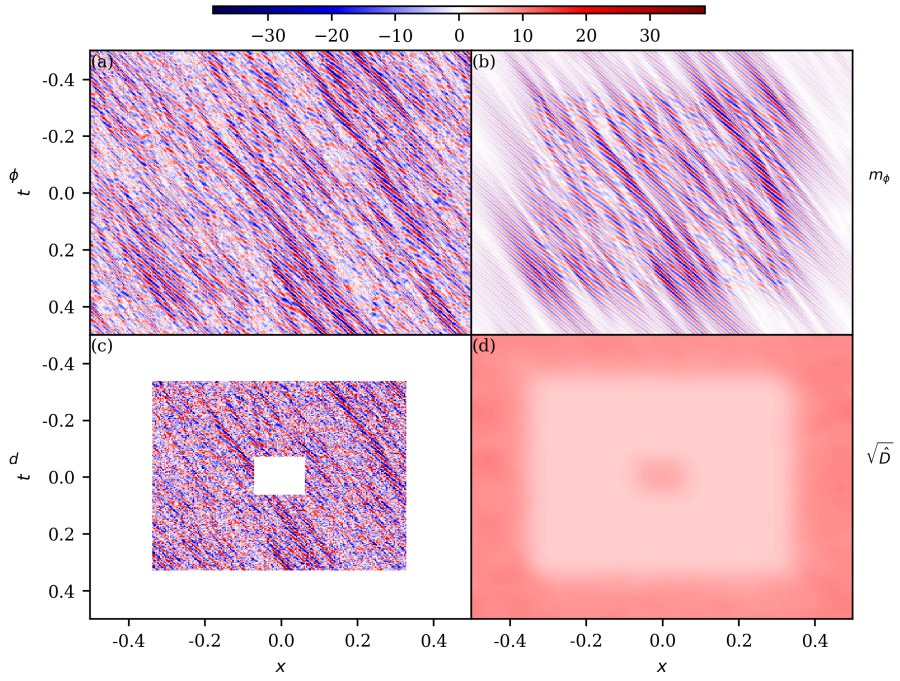

Figure 4: Signal ϕ ((a), drawn form the process corresponding to Eq. 37), resulting noisy measurement data d (c), reconstruction m_ϕ (b) and uncertainty map $\sqrt{\hat{D}}$ (d) of the reconstruction.

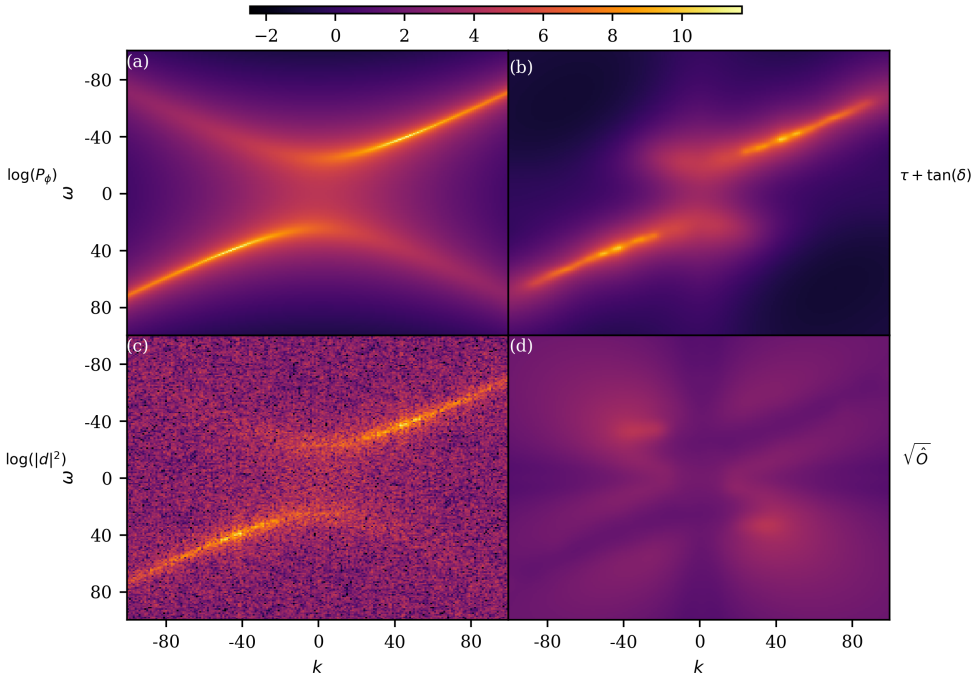

Figure 5: For the field and data shown in figure 4, natural logarithmic spectrum $\log(P_\phi)$ (a), projected data $\log\left(|d|^2\right)$ (c), reconstruction $\tau + \tan(\delta)$ (b) and uncertainty estimate $\sqrt{\hat{O}}$ (d) as defined in section 3.e.

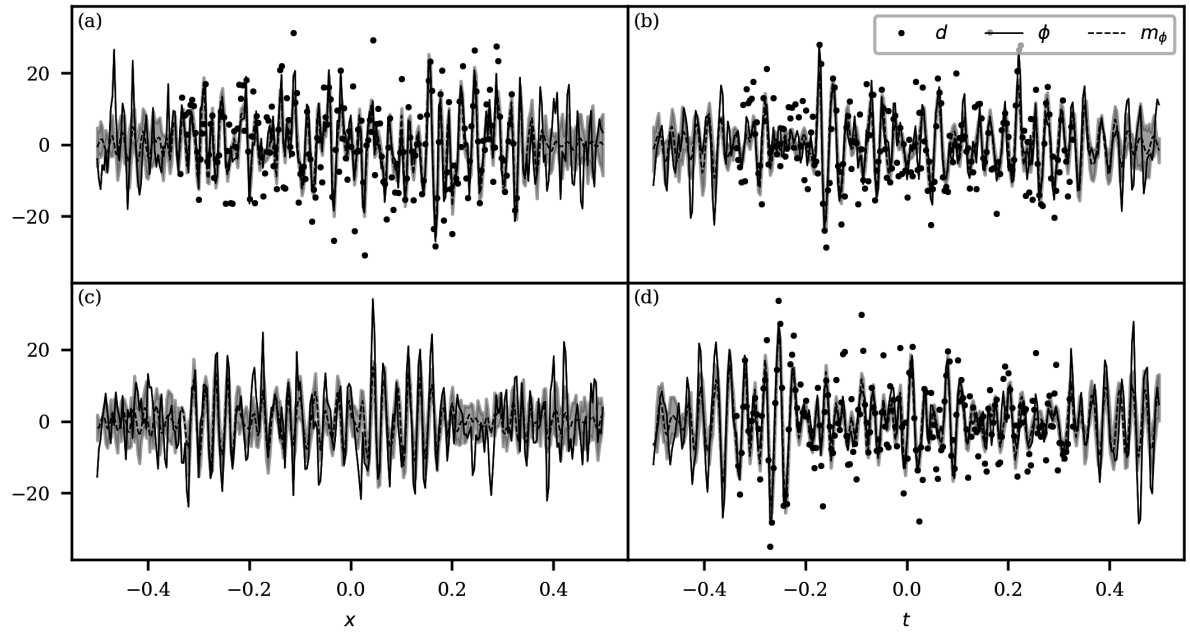

Figure 6: Slices through the full field, the data and the reconstruction, shown in figure 4. The left panels show the spatial structure for $t = -0.27$ (a) and $t = 0.38$ (c). The right panels show the temporal evolution at $x = -0.15$ (b) and at $x = 0.17$ (d). The gray areas denote the corresponding one-sigma uncertainty estimate.

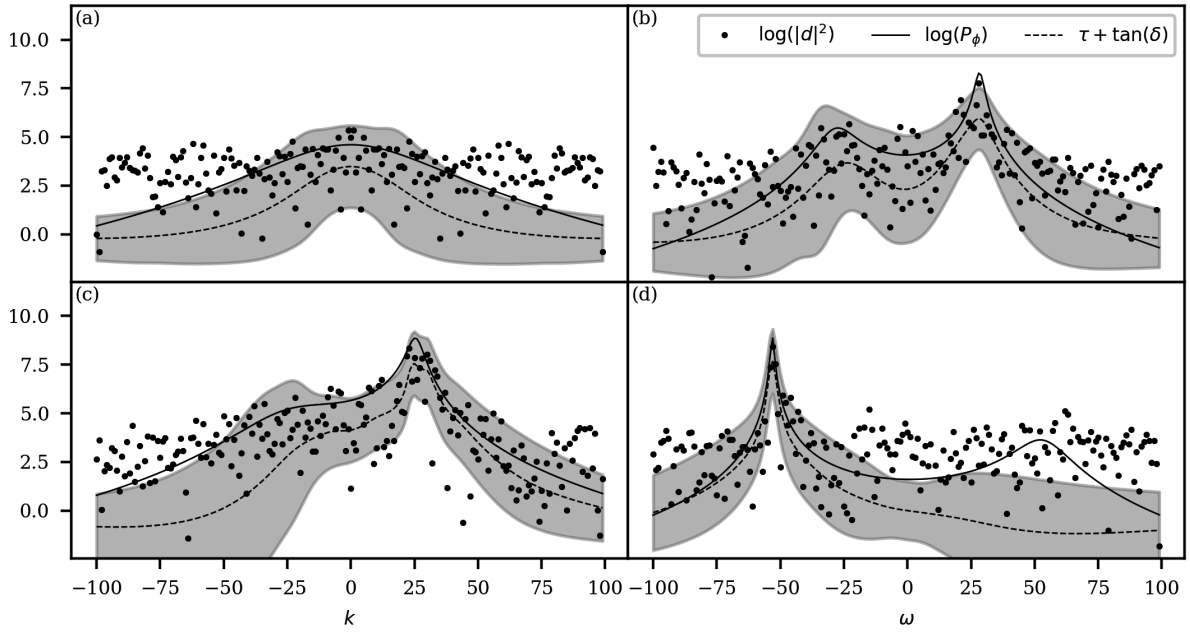

Figure 7: Slices through the natural logarithmic spectrum, the data and the reconstruction, shown in figure 5. The left panels show the spectrum as a function of k for fixed $\omega = 0$ (a) and $\omega = -30$ (c). The right panels show the spectrum as a function of ω for fixed $k = -20$ (b) and $k = 70$ (d). The gray areas denote the corresponding one-sigma uncertainty estimate.

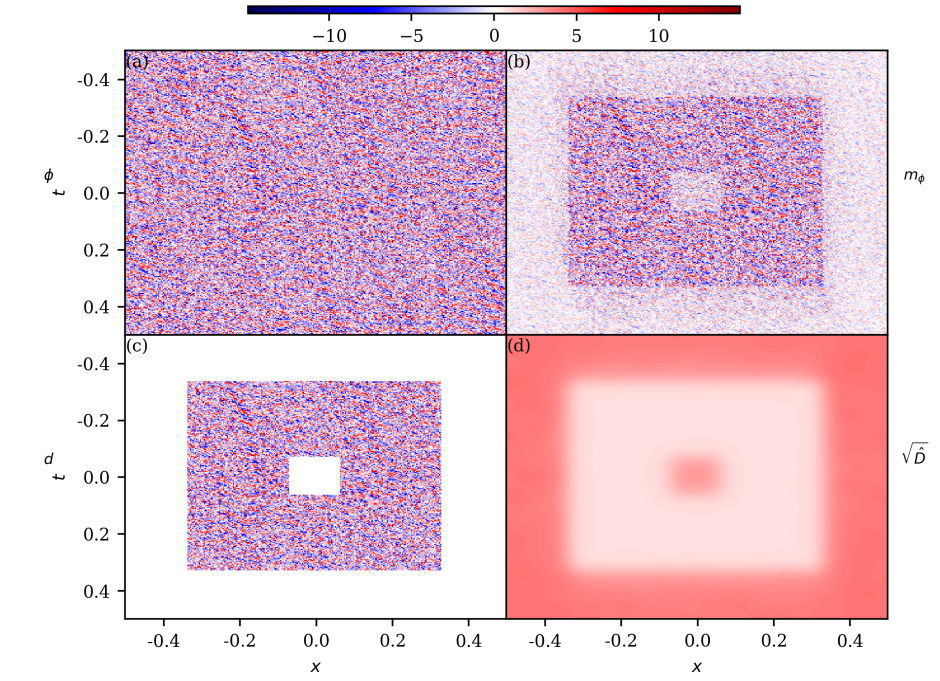

Figure 8: Signal ϕ (a), noisy measurement data d (c), reconstruction m_ϕ (b), and uncertainty map $\sqrt{\hat{D}}$ (d), for the highly structured spectrum given by Eq. 40

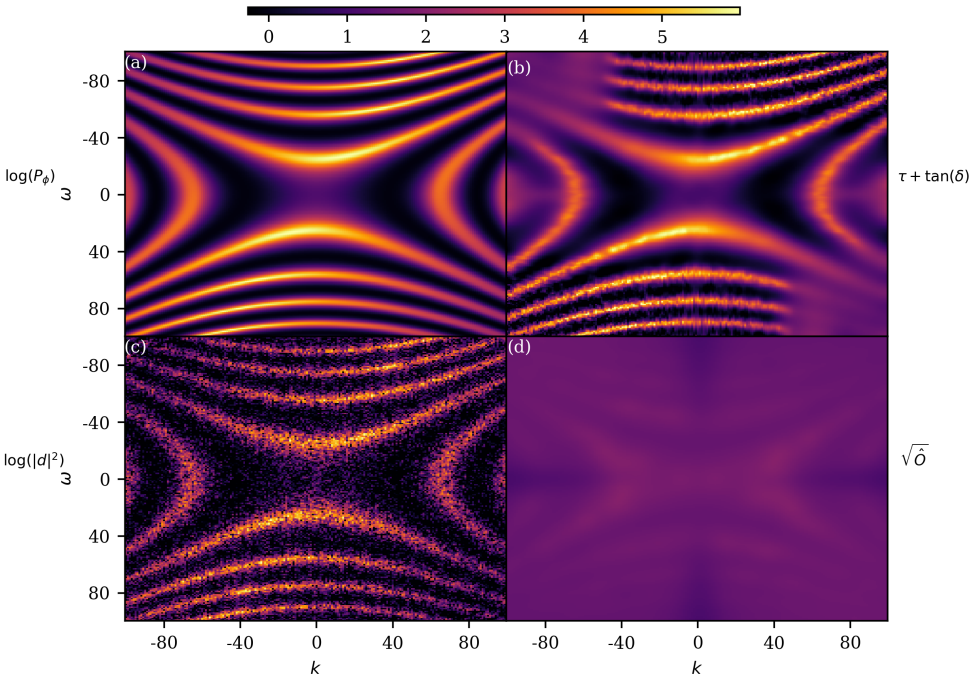

Figure 9: For the field and data shown in figure 8, natural logarithmic spectrum $\log(P_\phi)$ (a), projected data $\log\left(|d|^2\right)$ (c), reconstruction $\tau + \tan(\delta)$ (b) and uncertainty estimate $\sqrt{\hat{O}}$ (d).

5 Preliminary application to the Ionosphere

Spectral density estimation can be used to reconstruct fluctuations of the electron density in the ionosphere from radio observations. Ground based radio telescopes receive radio signals coming from the sky, however, these signals get distorted by the ionosphere before being received by the radio antennas. In general, both, the original astronomical signal as well as the ionospheric distortions are unknown. However, if one restricts the analysis to bright calibration sources, for which the flux-signal is known, one can use the resulting distortions in order to reconstruct fluctuations in the ionospheric electron density. These fluctuations are proportional to phase distortions of the received radio waves. We therefore aim to reconstruct the field of phase distortions arising from ionospheric distortions. To do so, we use calibration data of the LOw Frequency ARay (LOFAR [22]) stations. LOFAR is a set of phased array radio antennas, capable of detecting radio signals in the 10 - 240 MHz frequency range. The field of view of the antennas cover a large part of the sky above northern Europe, with a higher density in the Northern Netherlands (see figure 10).

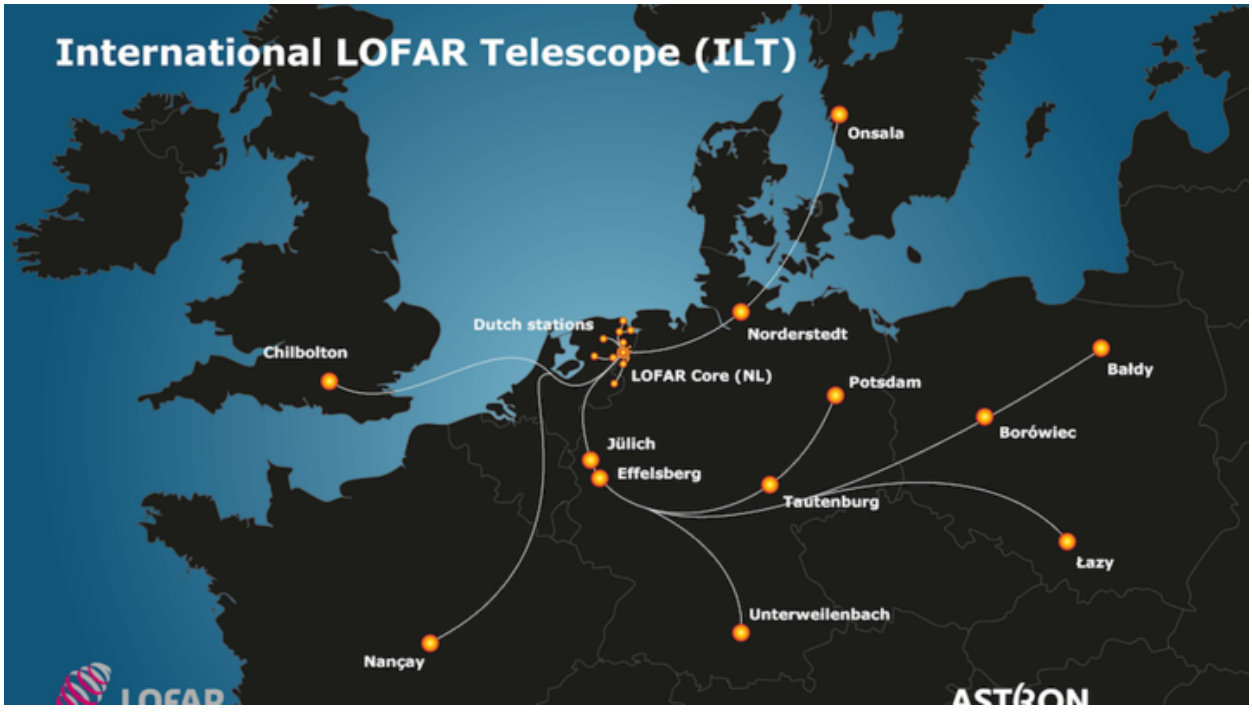

Figure 10: Locations of the LOFAR stations distributed over Europe[2]

[2]https://www.weltderphysik.de/gebiet/universum/teleskope-und-satelliten/radioteleskop-lofar/

5.a The Ionosphere

The ionosphere is defined as the ionized part of the Earth's atmosphere and is formed due to ionization of atoms and molecules by electromagnetic radiation and cosmic ray particles (see e.g. [26]). The ionized part of the atmosphere ranges from 50 to more than 500 km above the surface of the earth. The ionospheric electron density varies from 10^7 up to 10^{12} particles per cubic meter. It is typically described in terms of several layers: The innermost D-layer ranging from 50 to 90 km, the E-layer (90 - 150 km) and the outermost F-layer (150 - more than 500 km above the surface of the Earth). Ionization of the D- and E-layer is mostly caused by solar radiation, whereas the F-layer, the region with the highest concentration of free electrons, is additionally ionized due to cosmic radiation. Therefore the F-layer is often subdivided into two layers, the inner F1-layer corresponding to solar ionization and the outer F2-layer corresponding to ionization due to cosmic radiation [27]. Obviously we expect the properties of the ionosphere to differ during night and daytime since the D-, E- and F1-layer are only present during daytime whereas the F2-layer is present during day and nighttime.

The data used in this setup was measured around midnight and therefore the F2-layer mostly contributes to the phase distortions of radio waves. The properties of the electron density profile in this layer are caused by various different mechanisms ranging from solar activity to the Earths global wind system.

Due to the coupling of electrons in the ionosphere to the Earths magnetic field, the spatial distribution of the electron density becomes anisotropic. In addition, geomagnetic storms, for example caused by a solar wind shocks, can cause non-linear effects for which the ionospheric layers become unstable.

As the F2-layer is typically located within the Earth's thermosphere, the total electron density is also subject to atmospheric weather effects. Particularly winds in the thermosphere caused by temperature differences introduce anisotropic spatial distributions.

As we can see, the total distribution of electrons in the ionosphere is subject to several complicated and partially non-linear effects. However, particularly on smaller scales and time periods much smaller than a day, many important properties of the electron density can be captured within a linear, anisotropic model.

5.b Setup

In this preliminary approach we restrict ourselves to one frequency channel ($\nu = 120$ MHz) of the LOFAR core stations, which are a set of 48 stations located around Exloo in the Northern Netherlands. These stations are located within several square kilometers. Therefore we can approximate the relevant part of the surface of the Earth to be flat, which results in a two dimensional plane within which the stations are located. As indicated in figure 11 a), a station receives a signal from a calibration source located in the sky above,

from which data about the phase distortions can be calculated. In order to reconstruct distortions of the ionosphere, we have to project this data back to the ionosphere (see figure 11 b)). For the moment we describe the ionosphere as a two dimensional plane located at a fixed height above the surface of the Earth (100 km in this proof of concept setup[3]). This clearly is a strong limitation, as the ionosphere is actually an extended three dimensional object. We aim to alleviate this limitation in the future.

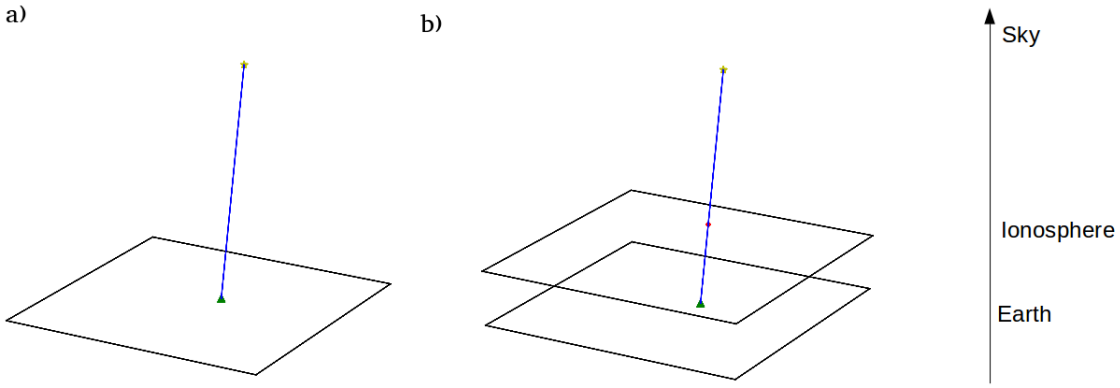

Figure 11: (a) Signal received from a LOFAR station for a calibration source. (b) Data back-projected to the ionospheric plane.

The LOFAR stations are capable of tracking multiple calibration sources simultaneously (42 in total in this setup). Therefore, for each station we get 42 different station-source pairs as well as corresponding data points (see figure 12). Ultimately, we use data for various moments in time (see figure 13), which enables us to infer properties of the dynamic evolution of the ionosphere from this dataset.

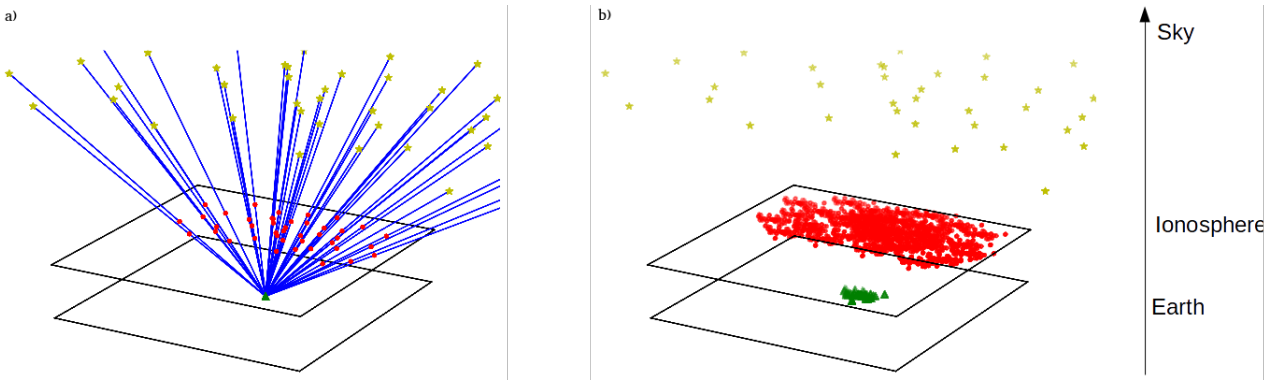

Figure 12: (a) Data received from a LOFAR station for several calibration sources, back-projected to the ionospheric plane. (b) Projected data from all LOFAR core stations and all sources for one moment in time.

[3]A more appropriate value from the International Reference Ionosphere (IRI [28]) model should be used for scientific results

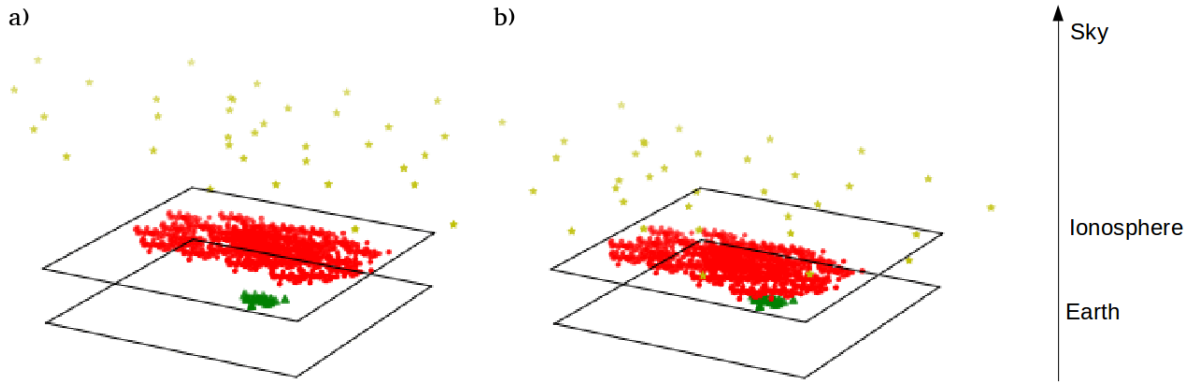

Figure 13: (a),(b): Projected data for different moments in time. The sources move along as the Earth rotates.

5.c Application

In the ionospheric plane we ultimately can define our field of interest: The 2+1 dimensional field of fluctuations associated with ionospheric distortions. We aim to reconstruct this field together with its spectral density, which also becomes a 2+1 dimensional field defined in the corresponding harmonic space.

Due to the fact that the projection between the ionospheric plane (called signal space in the future) and the LOFAR stations (called data space) is a linear operation, we can make use of a linear data model, as defined in Eq. 4. We also assume that the projected signal is subject to independent, additive Gaussian noise which models the uncertainty of the measured phase distortions provided by LOFAR.

We apply our method of spectral density estimation to this setup which results in reconstructions of the field of fluctuations, the corresponding spectrum, and uncertainty estimates for both fields. The results are discussed in the next section.

5.d Results

In figures 14 - 16 we depict snapshots of the spatial pattern of the reconstructed field, the projected data and corresponding uncertainties. As we can see, the algorithm reveals wave patterns, which seem to propagate as time passes. This becomes obvious if we look at the reconstruction of the spectral density displayed in figures 17 - 19, where we see that there exist dominating modes that follow a dispersion relation between $\mathbf{k}$ and ω, the coordinates of the harmonic partner domains of space and time, respectively.

In figure 20 we display the spatio-temporal evolution along one direction for the field of fluctuations, which clearly shows that spatial and temporal evolution is not independent. This becomes obvious if one considers the spectrum in the corresponding k-ω-plane, as shown in figure 21.

Furthermore, we also notice an odd behaviour of the system, namely that there exist structures which travel in the same manner as the data moves along the ionospheric

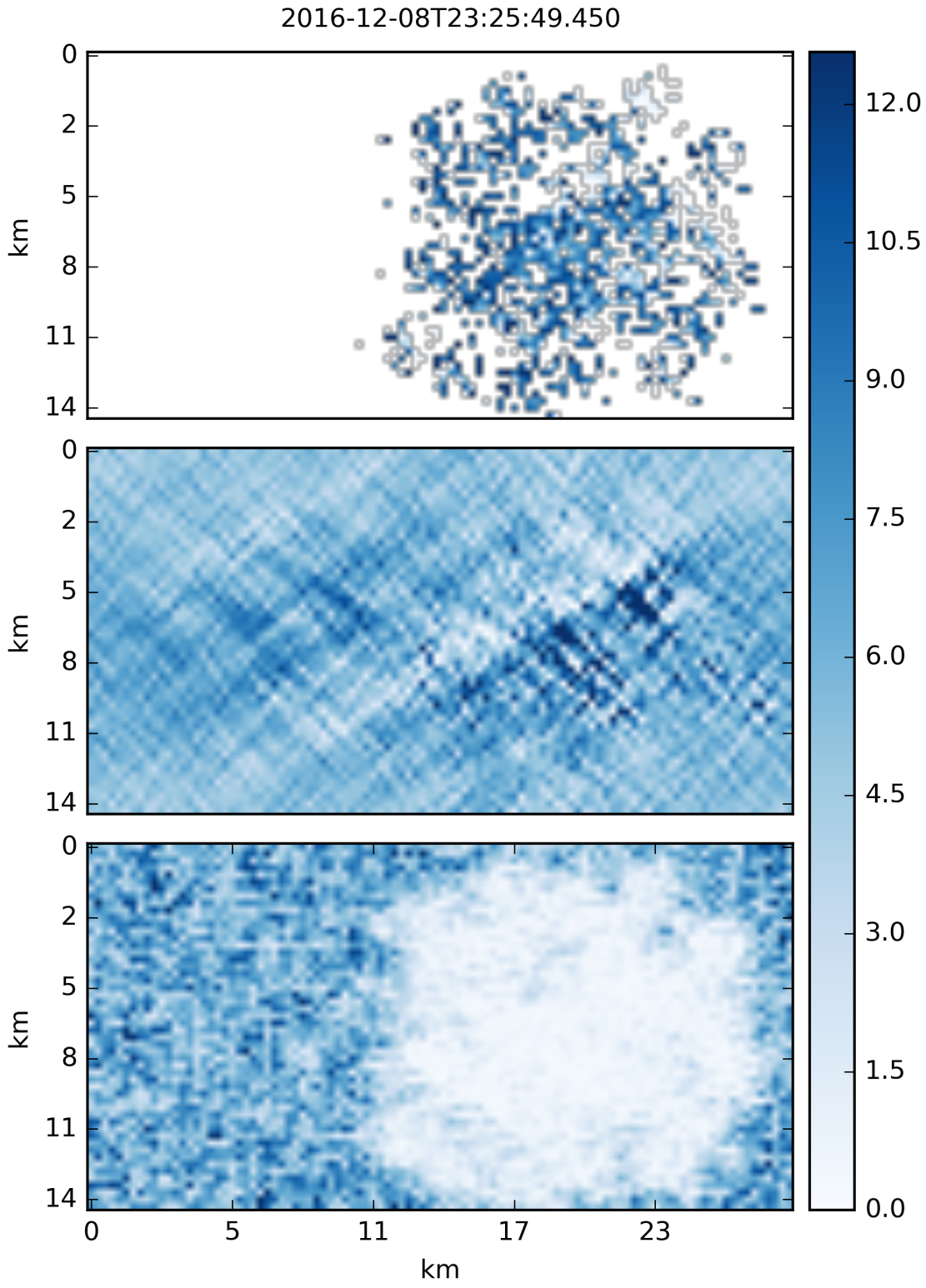

Figure 14: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the ionospheric plane for one moment in time. We see that the algorithm reconstructs a spatial pattern and partially extrapolates this pattern even to regions with no data. In addition the uncertainty is small in the patch where the data is located and is larger in regions of no data, as expected.

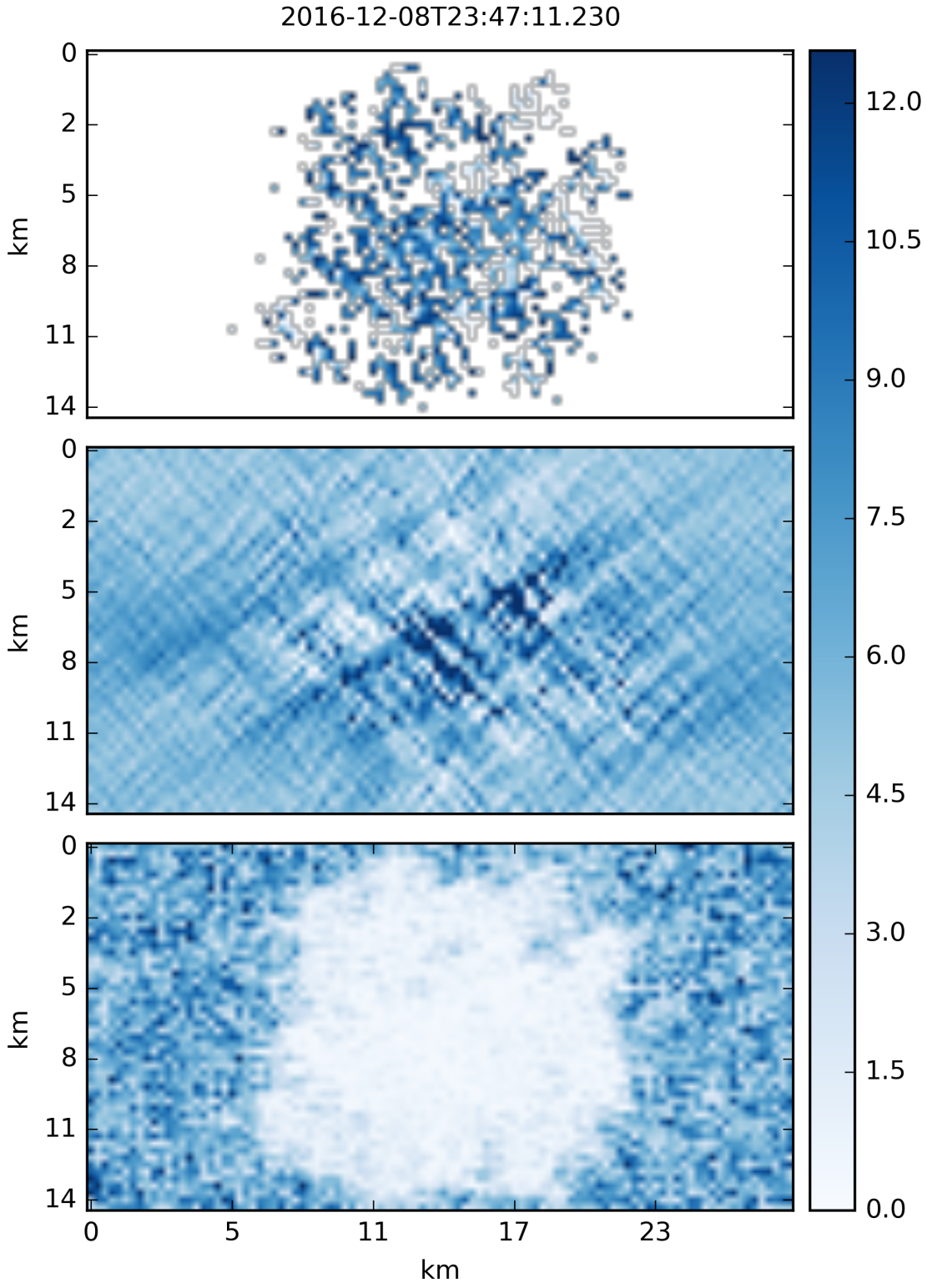

Figure 15: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the ionospheric plane for a later moment in time compared to the previous figure.

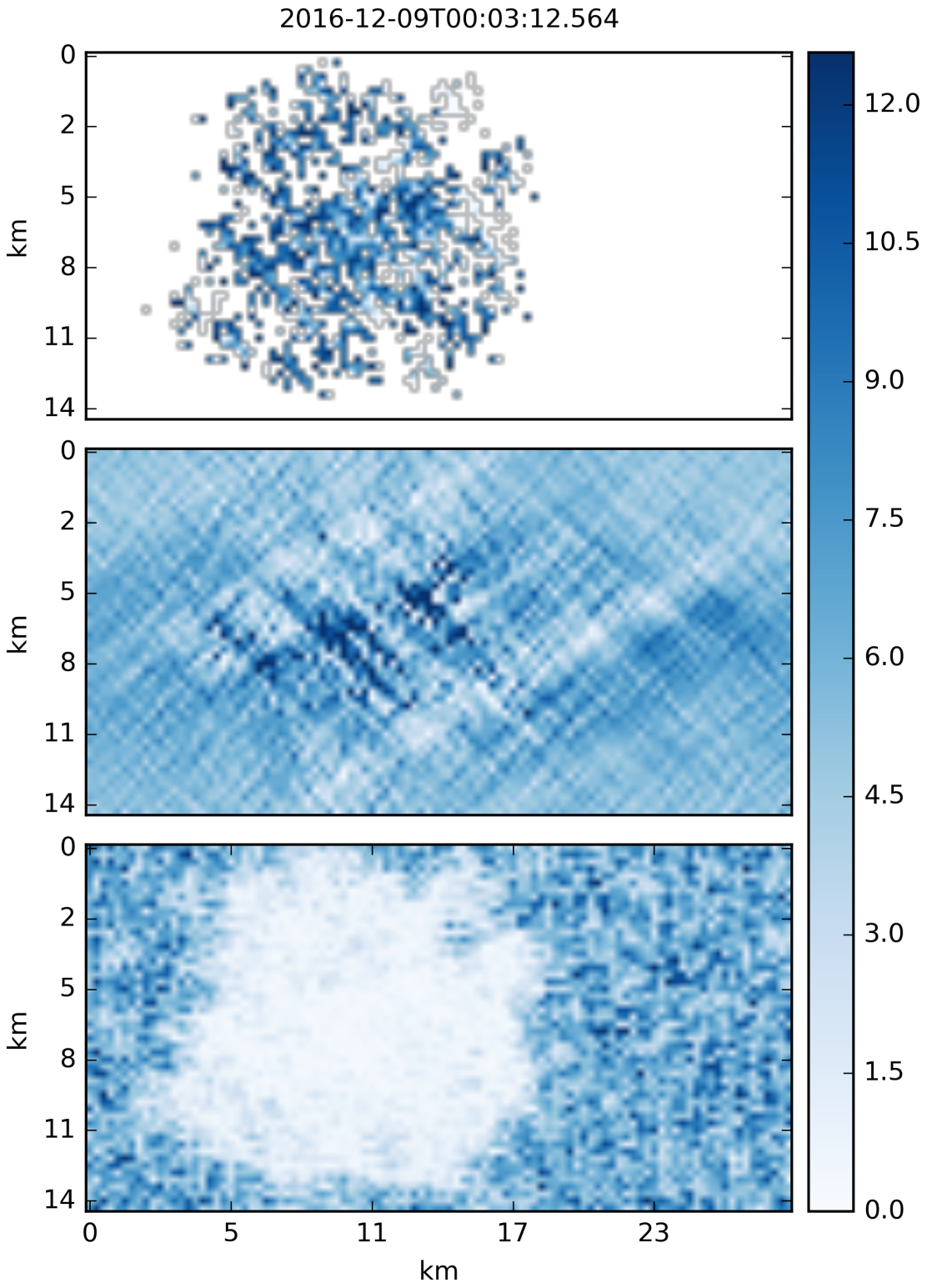

Figure 16: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the ionospheric plane for a later moment in time compared to the previous figure.

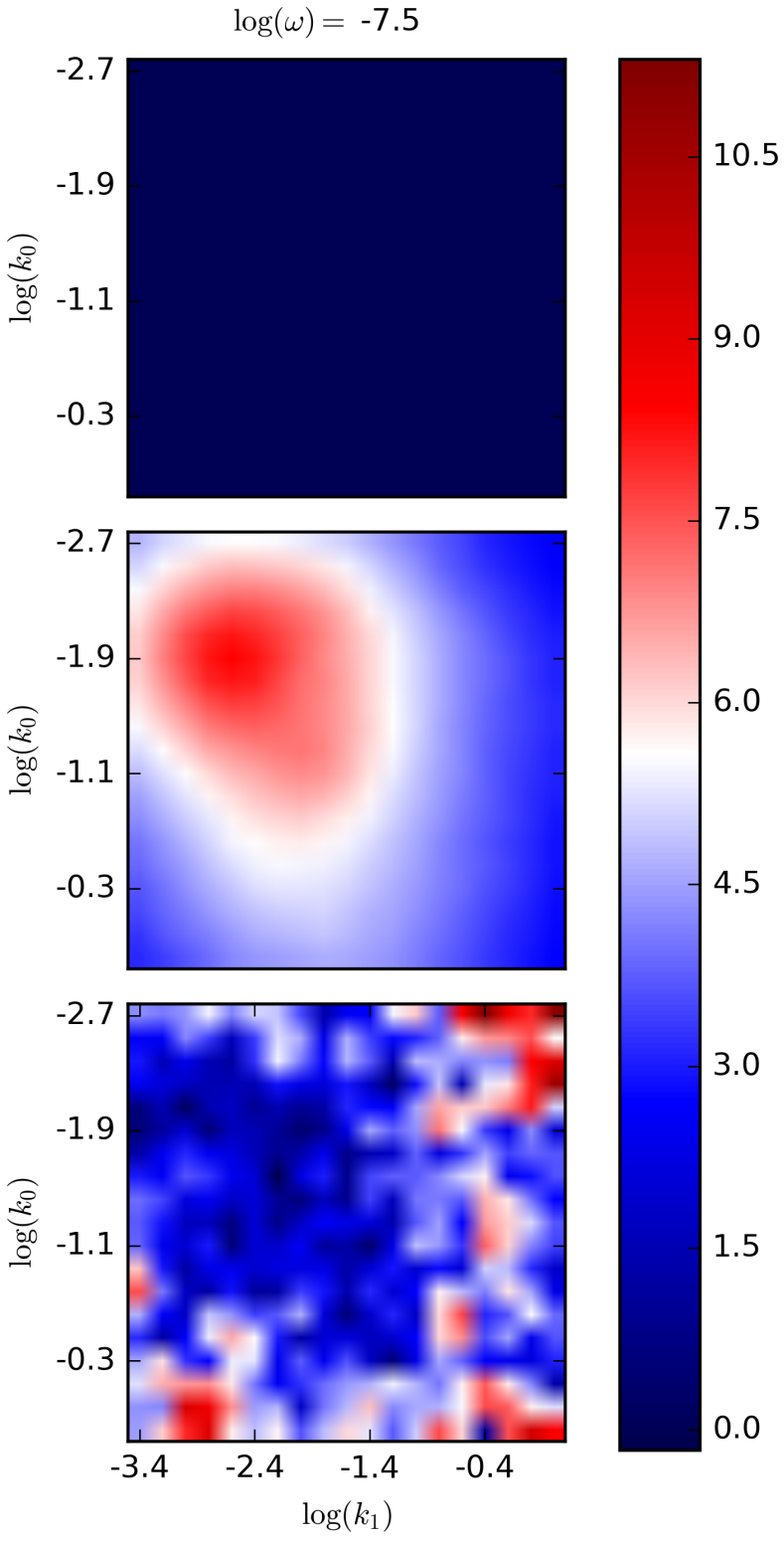

Figure 17: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale. We depict the spectrum in the log harmonic coordinates k_0, k_1 corresponding to the spatial coordinates, for a fixed ω, which is te coordinate in the harmonic partner domain of the temporal domain.

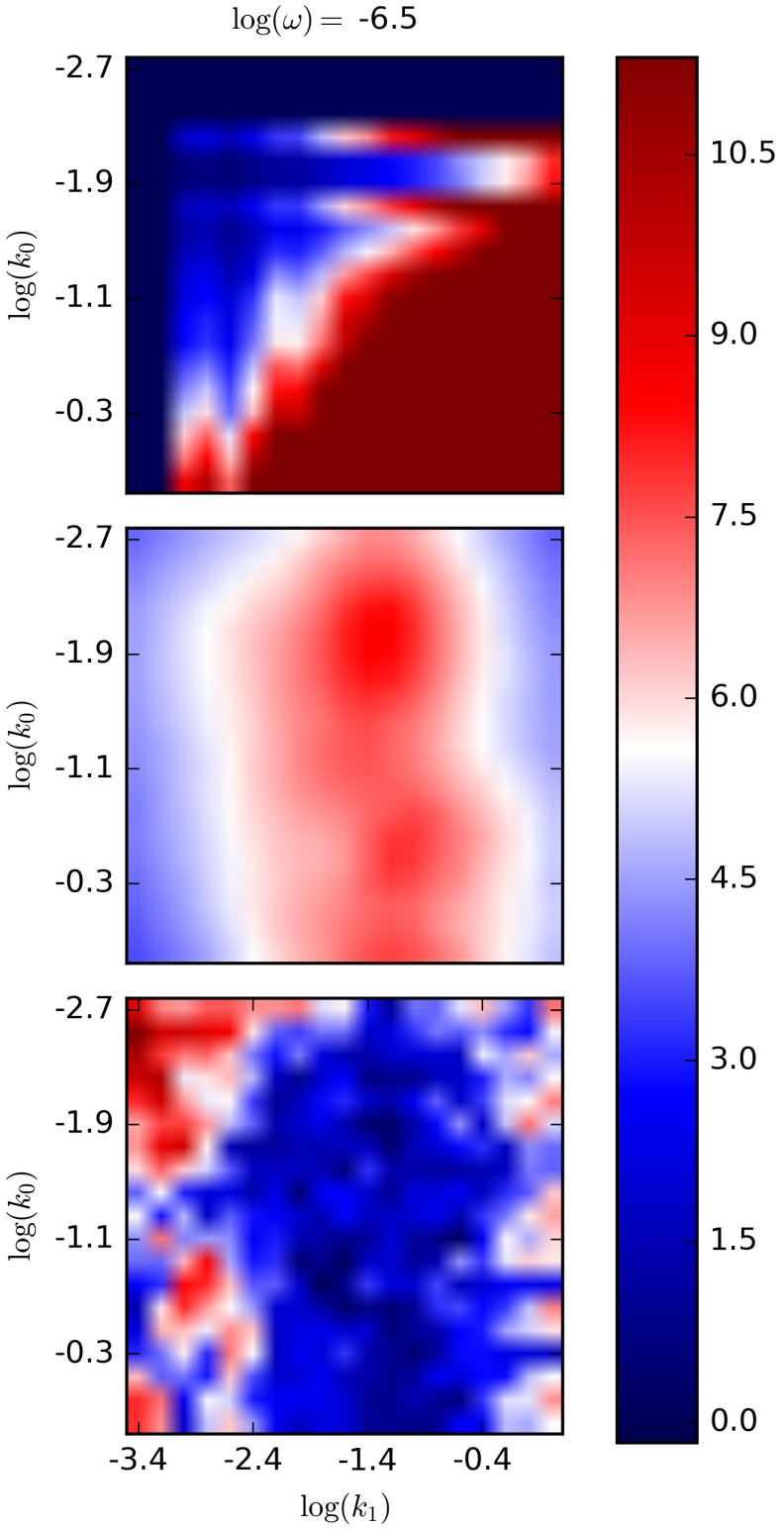

Figure 18: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for a larger ω compared to the precious figure.

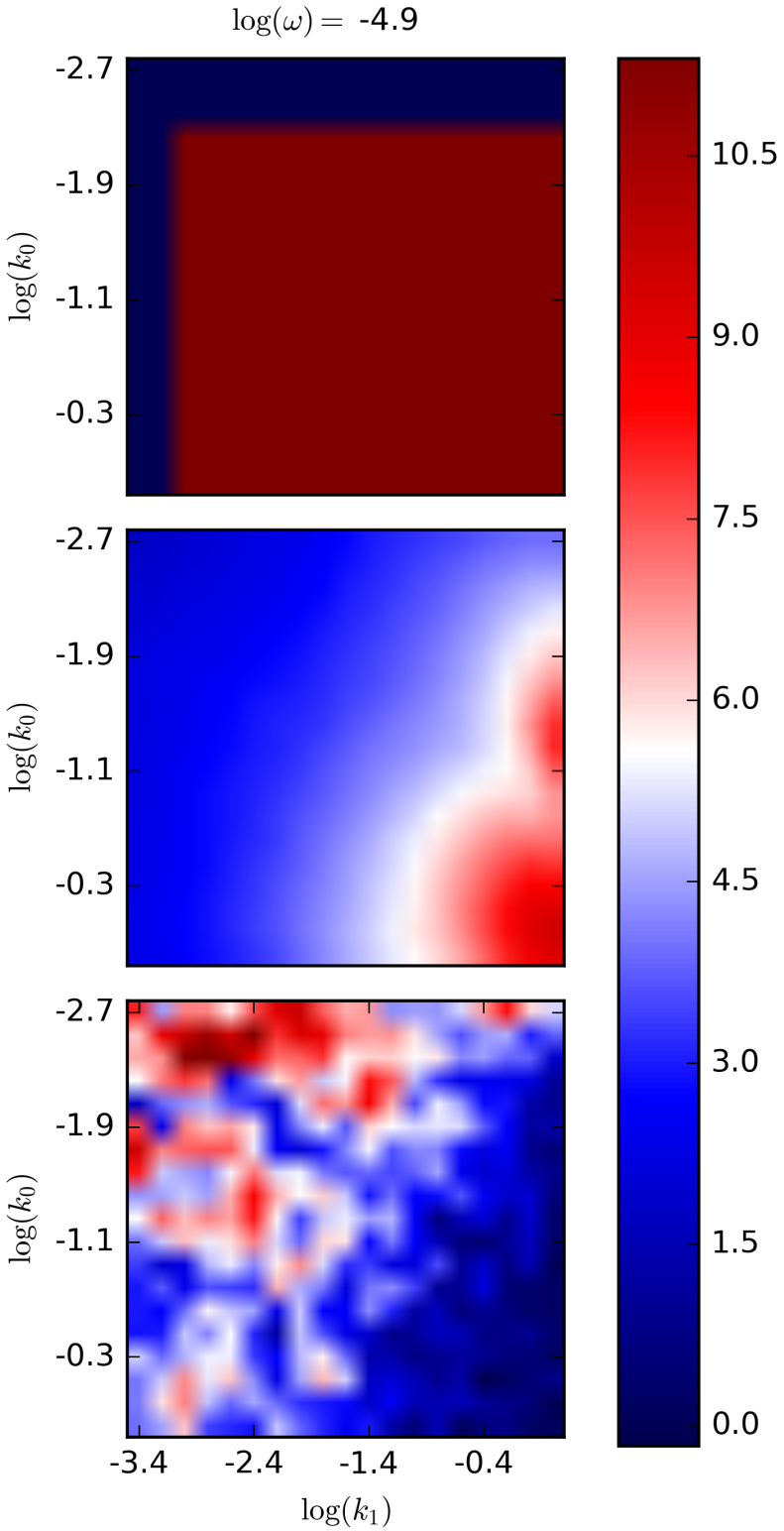

Figure 19: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for a larger ω compared to the precious figure.

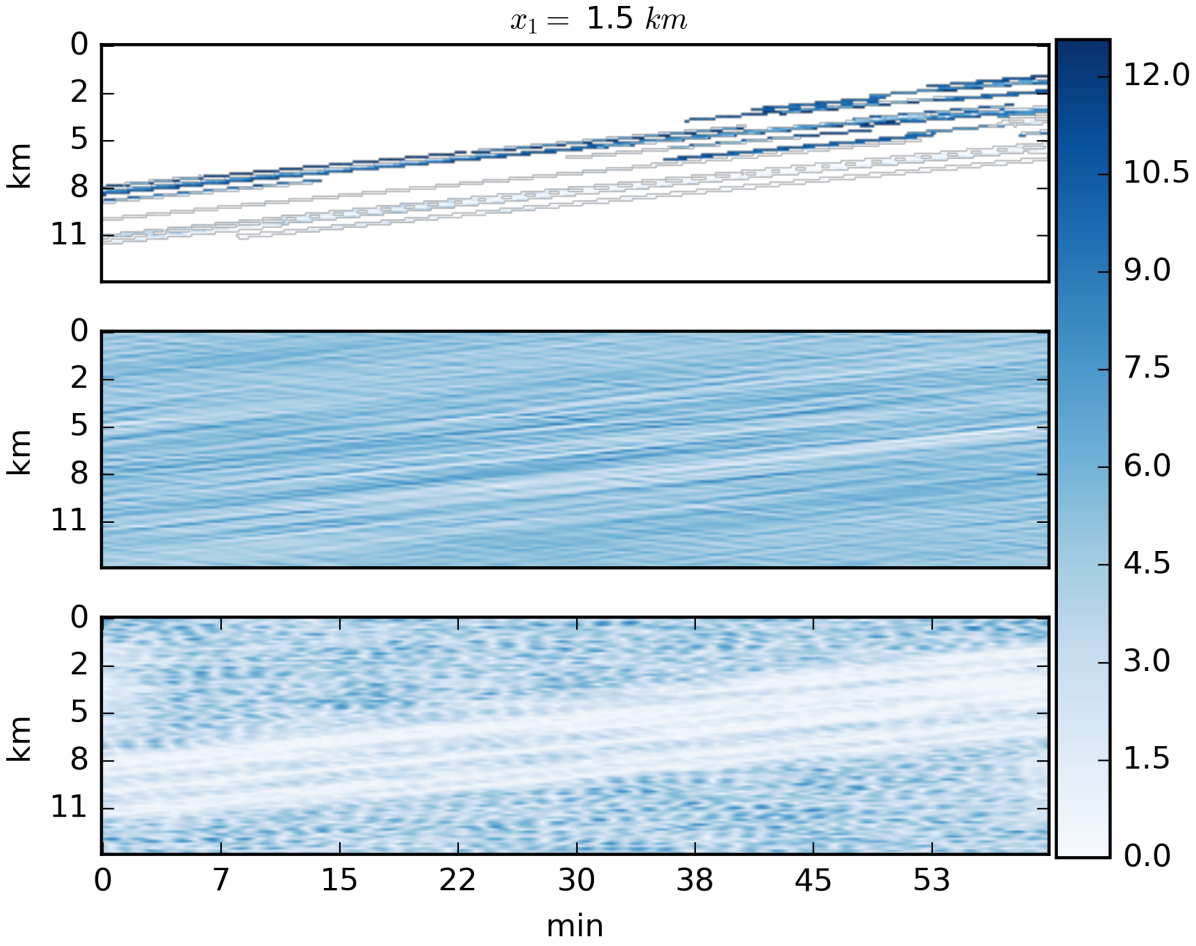

Figure 20: 1+1 dimensional spatio-temporal evolution of the data (top), the reconstruction (middle) and the corresponding one sigma uncertainty (bottom) for a fixed value of x_1, the other spatial coordinate. We see that the temporal evolution and the spatial patterns are not independent. Furthermore we notice that there exist dominant waves in the reconstruction which travel in the same way as the data moves along the ionospheric plane due to the rotation of the Earth.

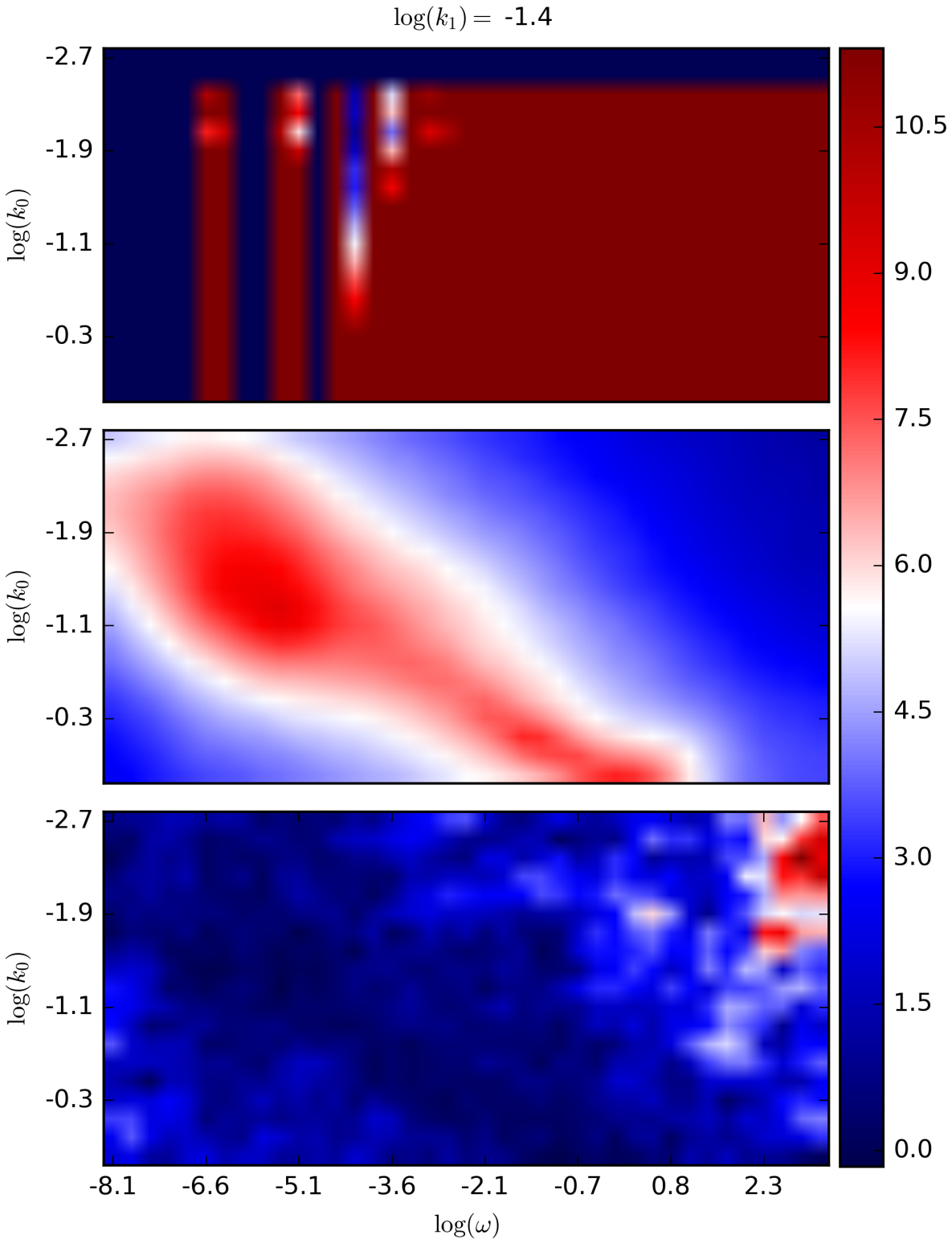

Figure 21: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the $\log(k_0)$-$\log(\omega)$-plane for a fixed $k1$. We notice that there is a strong correlation between the spectral modes of space and time. These modes are ultimately responsible for the dominant traveling waves seen in the previous figure.

plane. As the movement of the data is due to the rotation of the Earth, these structures correspond to traveling waves in the ionosphere with a velocity that matches the rotation velocity of the Earth. These waves appear to be non-physical, as there is no physical reason for an overpopulation of waves with this exact velocity. Consequently, we believe that there exist systematic errors in the data, which imprint these structures. These errors not only affect the reconstruction of the distortion field, but also the reconstruction of the spectrum due to the fact that the corresponding waves show a characteristic spectrum with a dispersion relation of the form

$$\omega = \mathbf{v} \cdot \mathbf{k} \ , \tag{41}$$

where $\mathbf{v}$ corresponds to the rotation velocity of the Earth at the location of the ionospheric plane. We aim to investigate these systematic effects in the following.

5.d.1 Implications of systematic effects

Lets assume that there is an overall positive offset in the calibration data for one station-source pair, compared to the other station-source pairs. As time passes, this data point moves along the ionospheric plane, together with the rest of the data. Due to the offset, the fluctuation data is always larger for this point compared to the others. This structure looks like a traveling wave front, and is interpreted as such by the algorithm due to the fact that our data model does not have a notion of systematic errors. Particularly, all source of uncertainty is assumed to be independent Gaussian noise which ultimately means that all dependent structures have to be explained by the reconstruction of the signal.

To overcome this issue we therefore would have to introduce such systematic errors in the data model, in order to properly deal with these effects. However, this requires prior knowledge about the origin of such systematic errors. Further investigations of the procedure how the distortion data is revealed from LOFAR measurements have to be made in order to define such models.

Luckily there exists another way of how to proceed in this context. As we aim to reconstruct a field of fluctuations, it is possible to pre-process the data by subtracting a mean from the data, without changing the statistics of the fluctuations. We therefore subtract the temporal mean from each time set of data points corresponding to one station-source pair (see again figure 11 a for an illustration of such a pair). If the source of systematic error is an offset in the data, which is constant in time as described above, this procedure is able to remove this systematics.

5.d.2 Subtracting the temporal mean

The processed data set d' takes the form

$$d'_{tij} = d_{tij} - \frac{1}{T}\int_0^T d_{t'ij}\ dt' \ , \tag{42}$$

where (t, i, j) are the temporal, station and direction indices respectively and d is the original data. Note that instead of producing a modified dataset d', this subtraction could also be introduced in a modified response operator, due to the fact that Eq. 42 is a linear operator acting on the old dataset d. This could provide a first step towards more sophisticated data models. However, in this preliminary application we used the processed dataset d' together with the old response operator instead of reformulating the data model.

We also applied our reconstruction algorithm to this dataset. The results for the reconstructed field of fluctuations are shown in figures 22 - 24 and the results for the spectrum in figures 25 - 27.

5.d.3 Discussion

Comparing the results for the original dataset d with the results from the processed data d' indicates that the proposed procedure is a good step towards the problem of dealing with these systematic effects. The spatial pattern of the reconstructed field of fluctuations shows less structures that travel with the rotation speed of the Earth. This preprocessing also partially removes the spectral parts associated with the dispersion relation corresponding to these systematic effects (Eq. 41). This becomes obvious if we compare figure 21 with figure 29, where we depict slices of the full spectra in the k-ω-plane.

Furthermore, if we consider the spatio-temporal evolution of the fluctuations for both cases, we see that the reconstruction corresponding to d' (figure 28) shows less structures matching the evolution of the data as the Earth rotates compared to the reconstruction from d (figure 20). This indicates that the reconstructions for d' reveal structures which are closer to physical ionospheric distortions. However, a more sophisticated treatment of the systematic effects in terms of a reformulated data model is necessary in order to fully remove these systematics which is beyond the scope of this work.

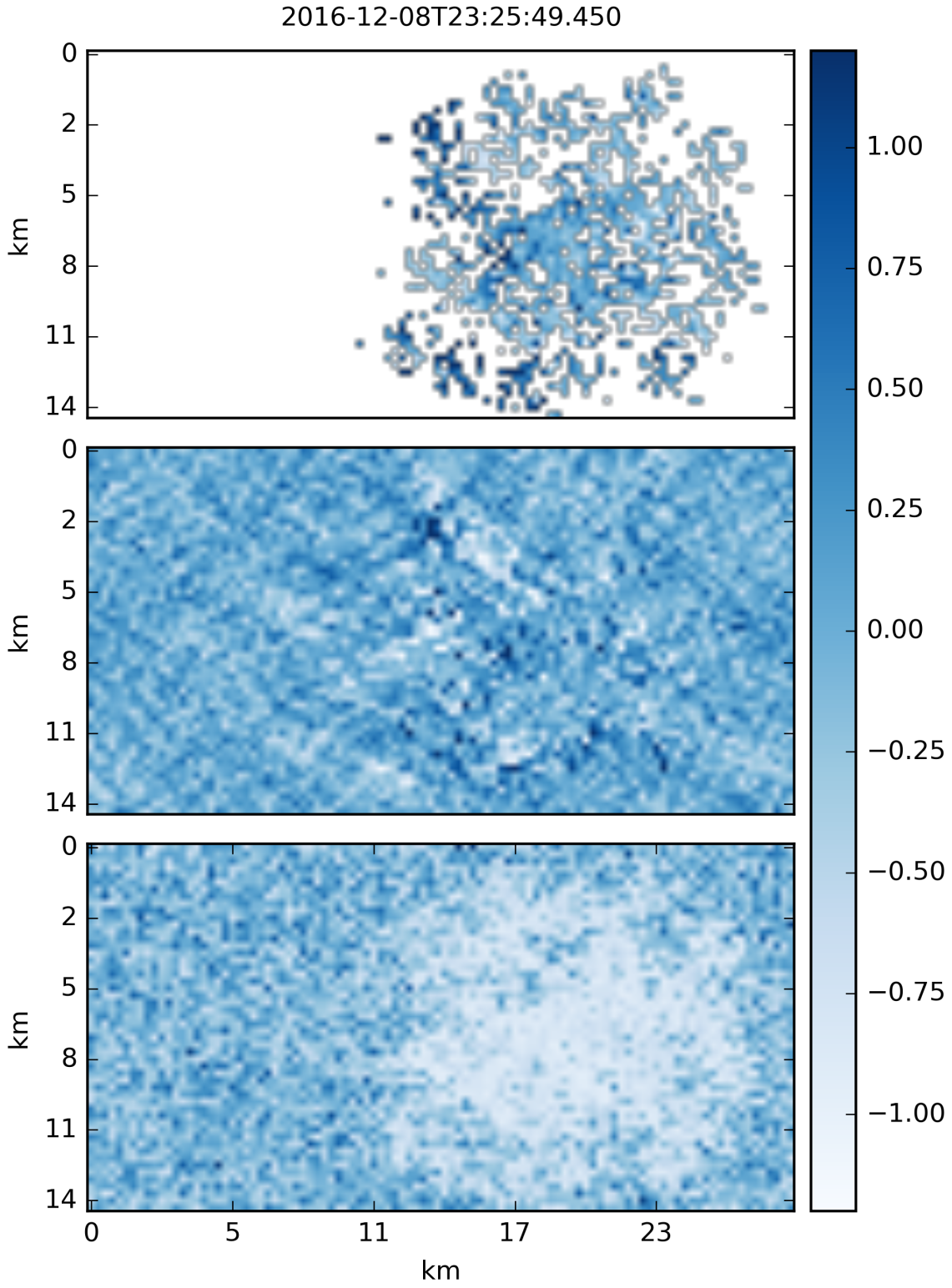

Figure 22: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the ionospheric plane for the processed dataset d' (Eq. 42). We notice that compared to the reconstruction using the original dataset d (see figures 14 - 16) there appear to be less structures imprinted due to systematic effects.

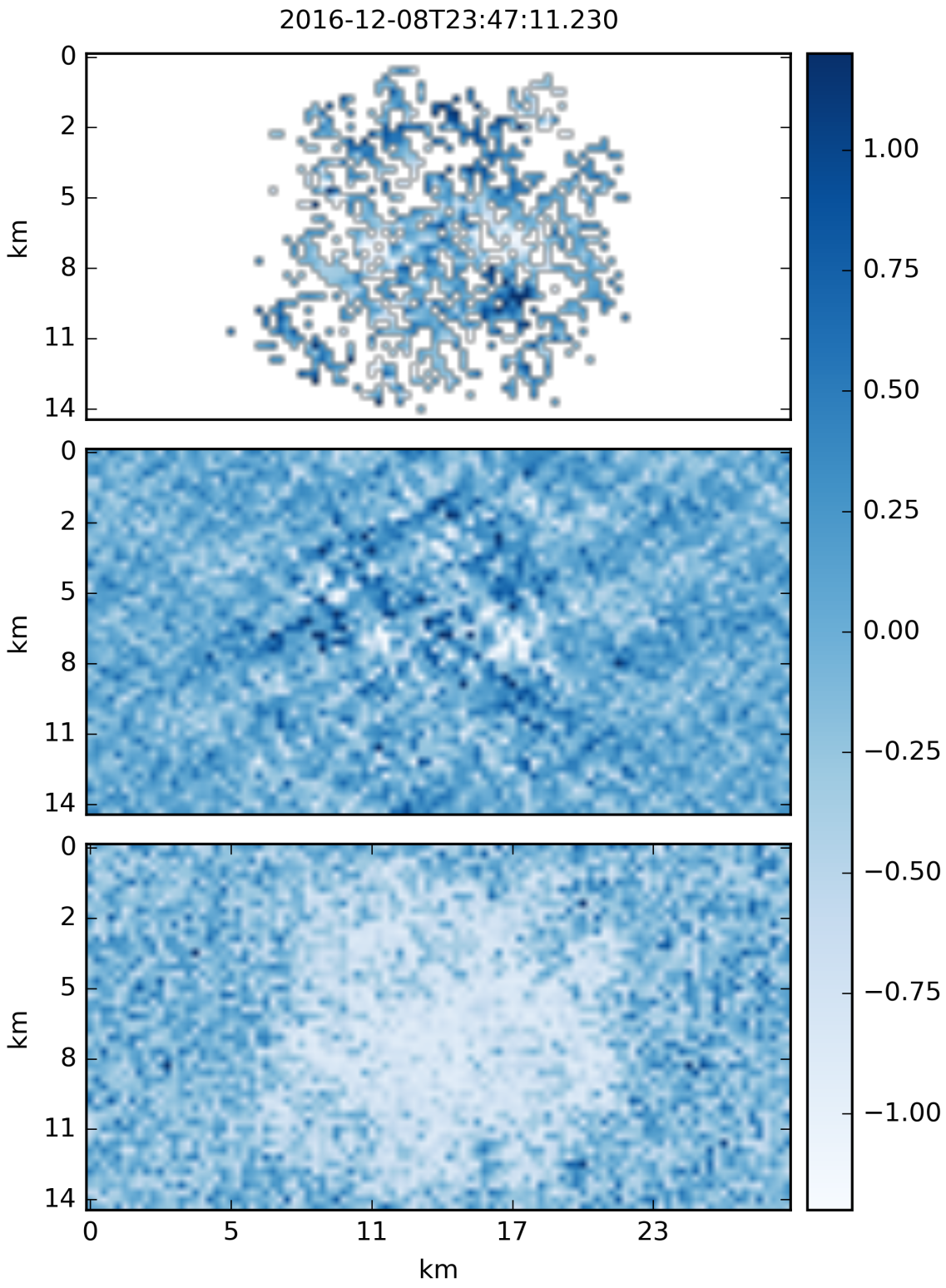

Figure 23: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the ionospheric plane for a later moment in time compared to the previous figure.

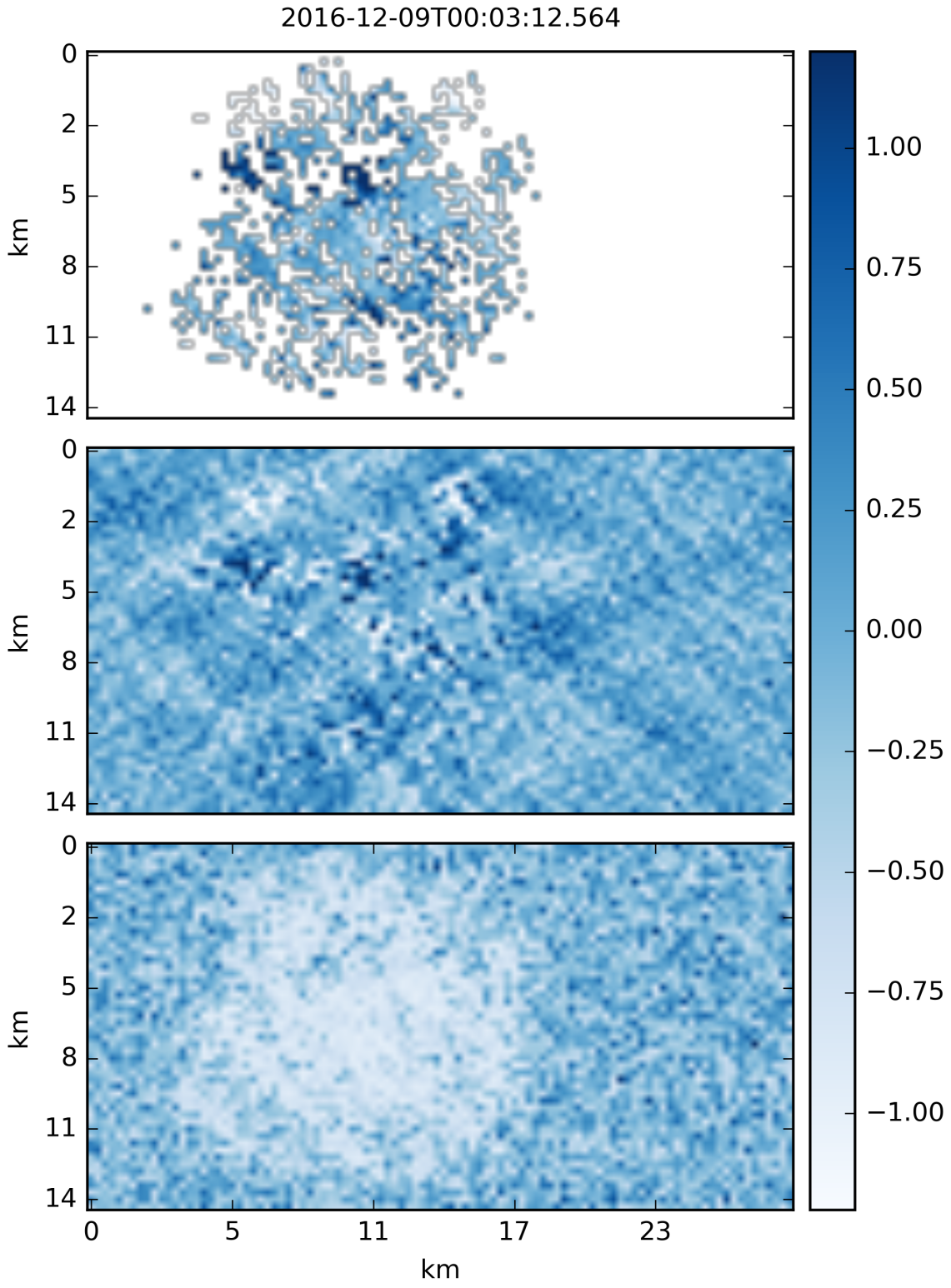

Figure 24: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the ionospheric plane for a later moment in time compared to the previous figure.

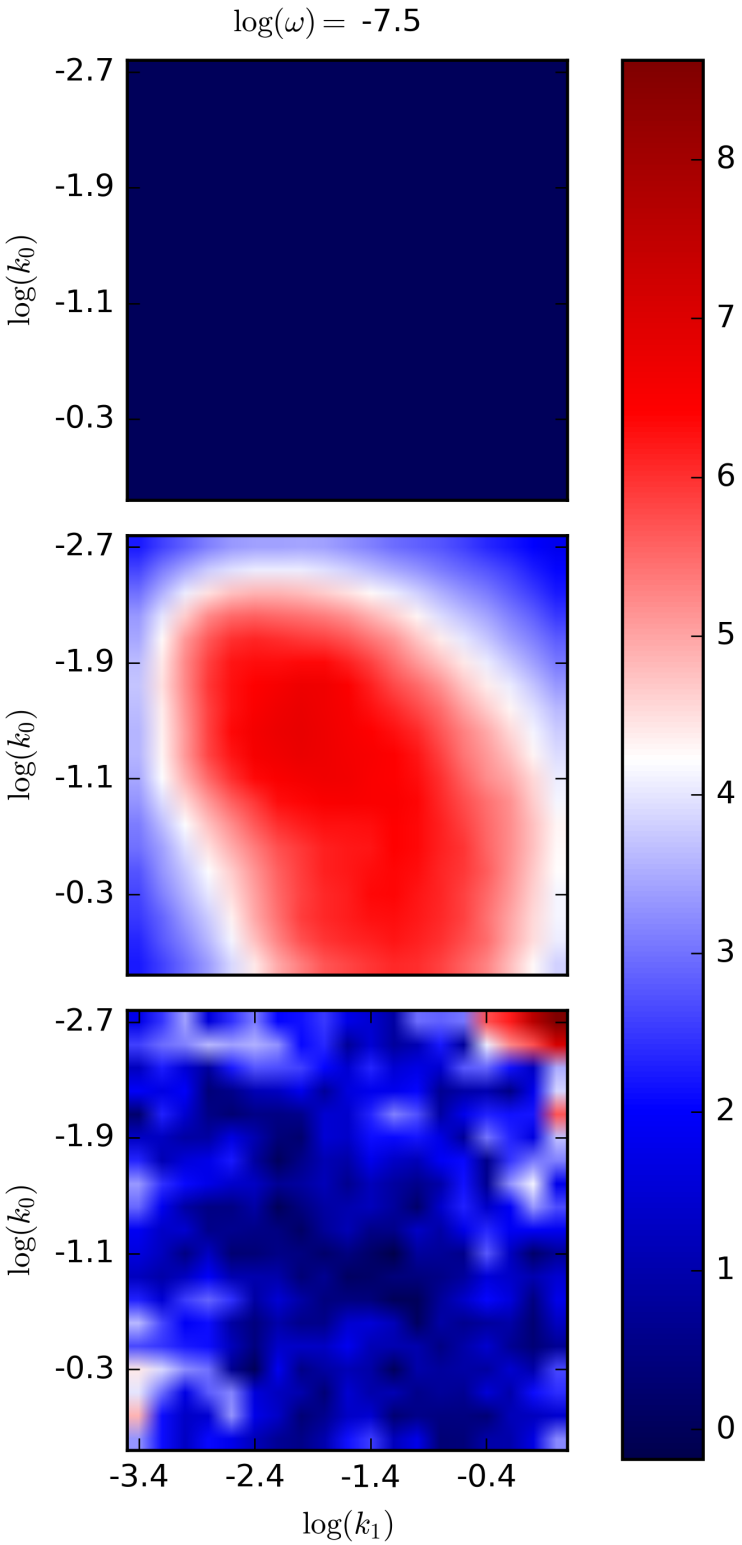

Figure 25: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for the processed dataset d' (Eq. 42).

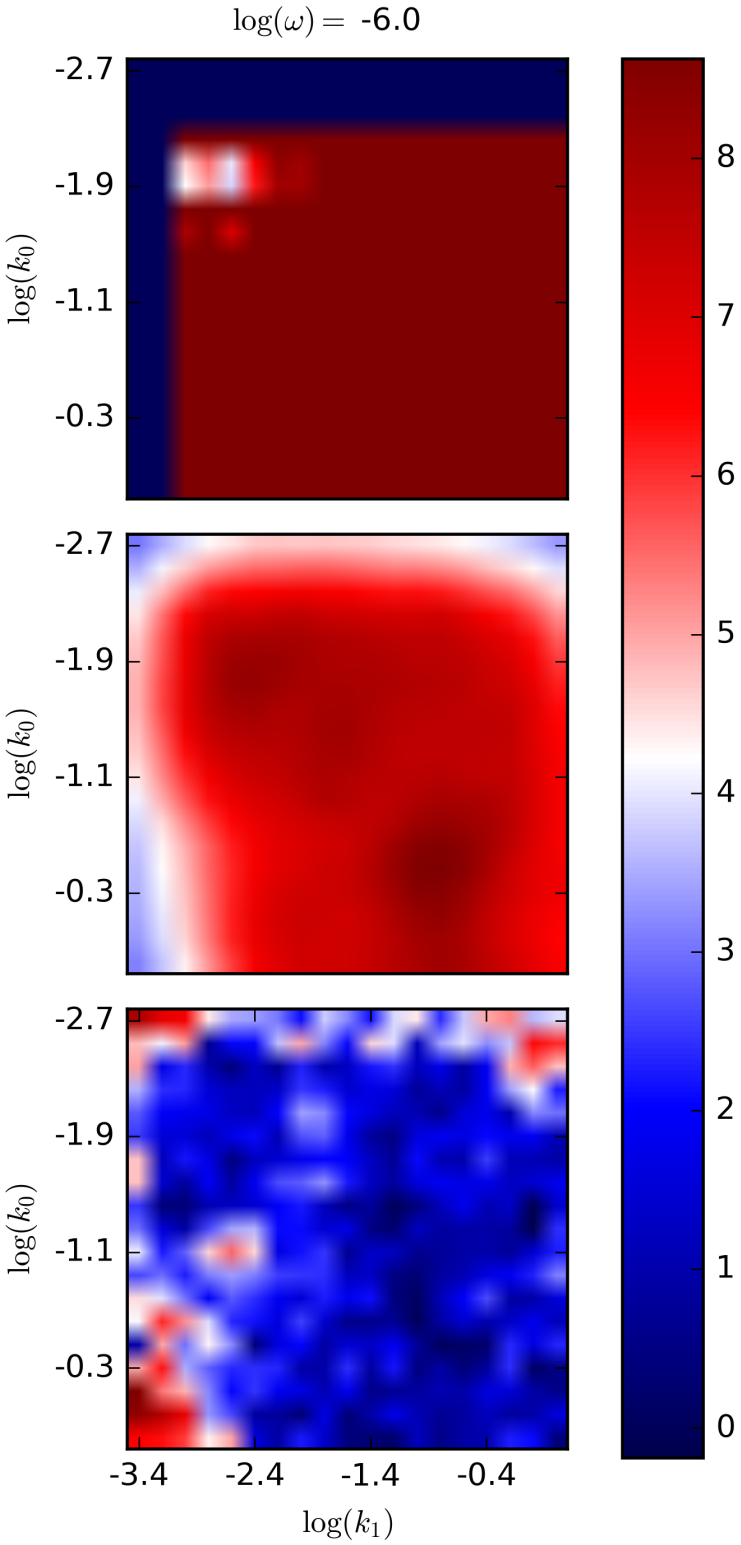

Figure 26: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for a larger ω compared to the previous figure.

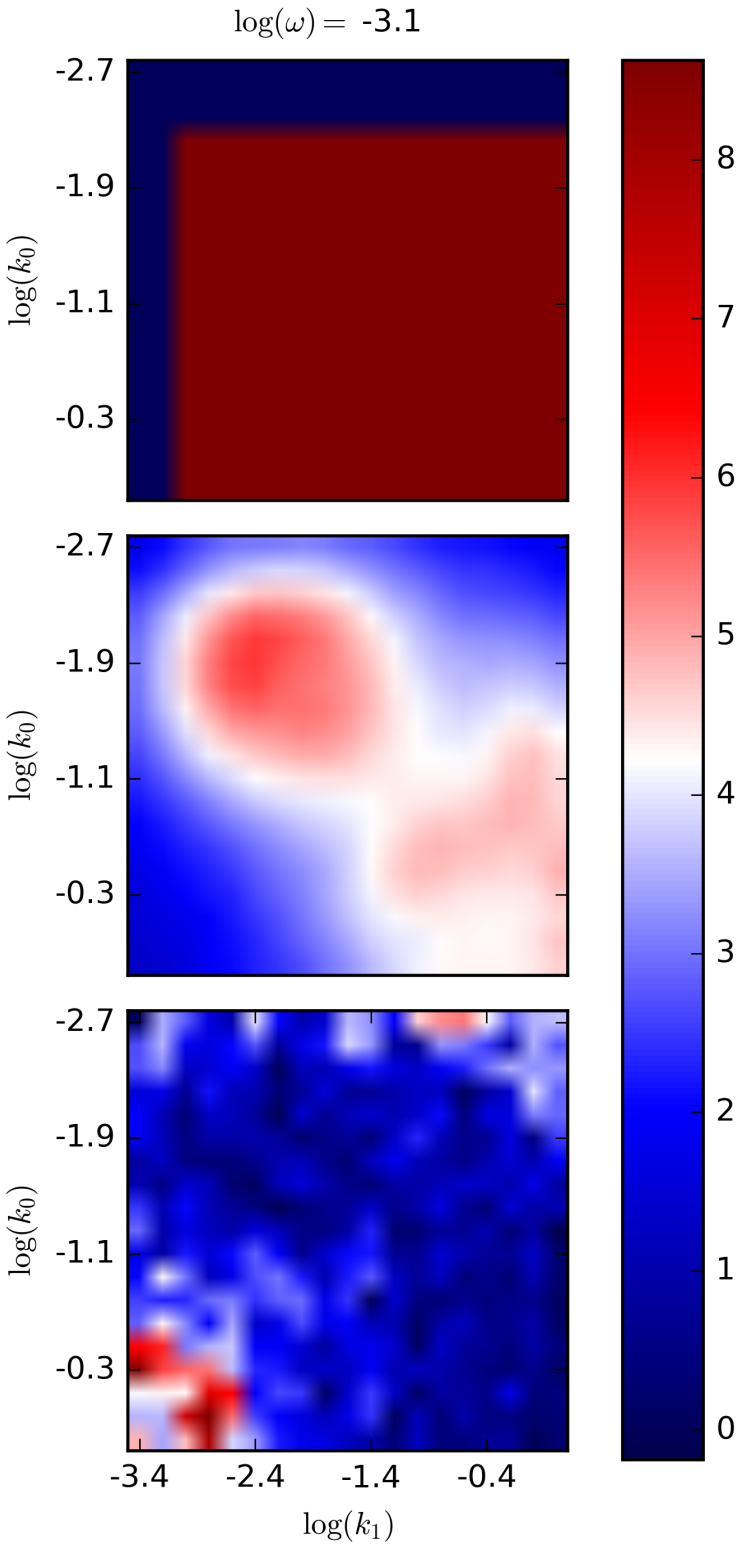

Figure 27: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for a larger ω compared to the previous figure.

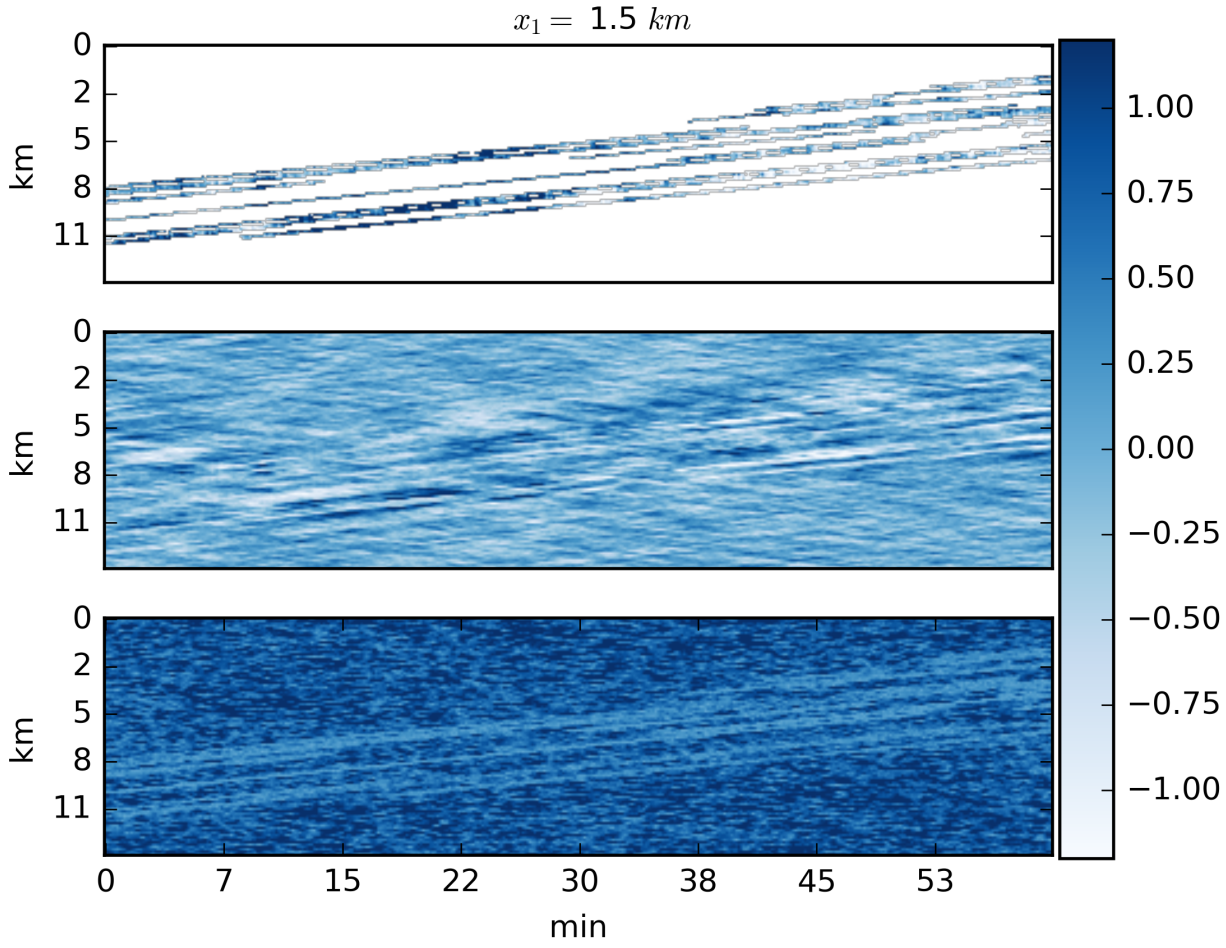

Figure 28: 1+1 dimensional spatio-temporal evolution of the data (top), the reconstruction (middle) and the corresponding one sigma uncertainty (bottom) for a fixed value of x_1, the other spatial coordinate. We see that the temporal evolution and the spatial patterns are not independent. However, there exist less structures that coincide with the rotation of the Earth compared to figure 20.

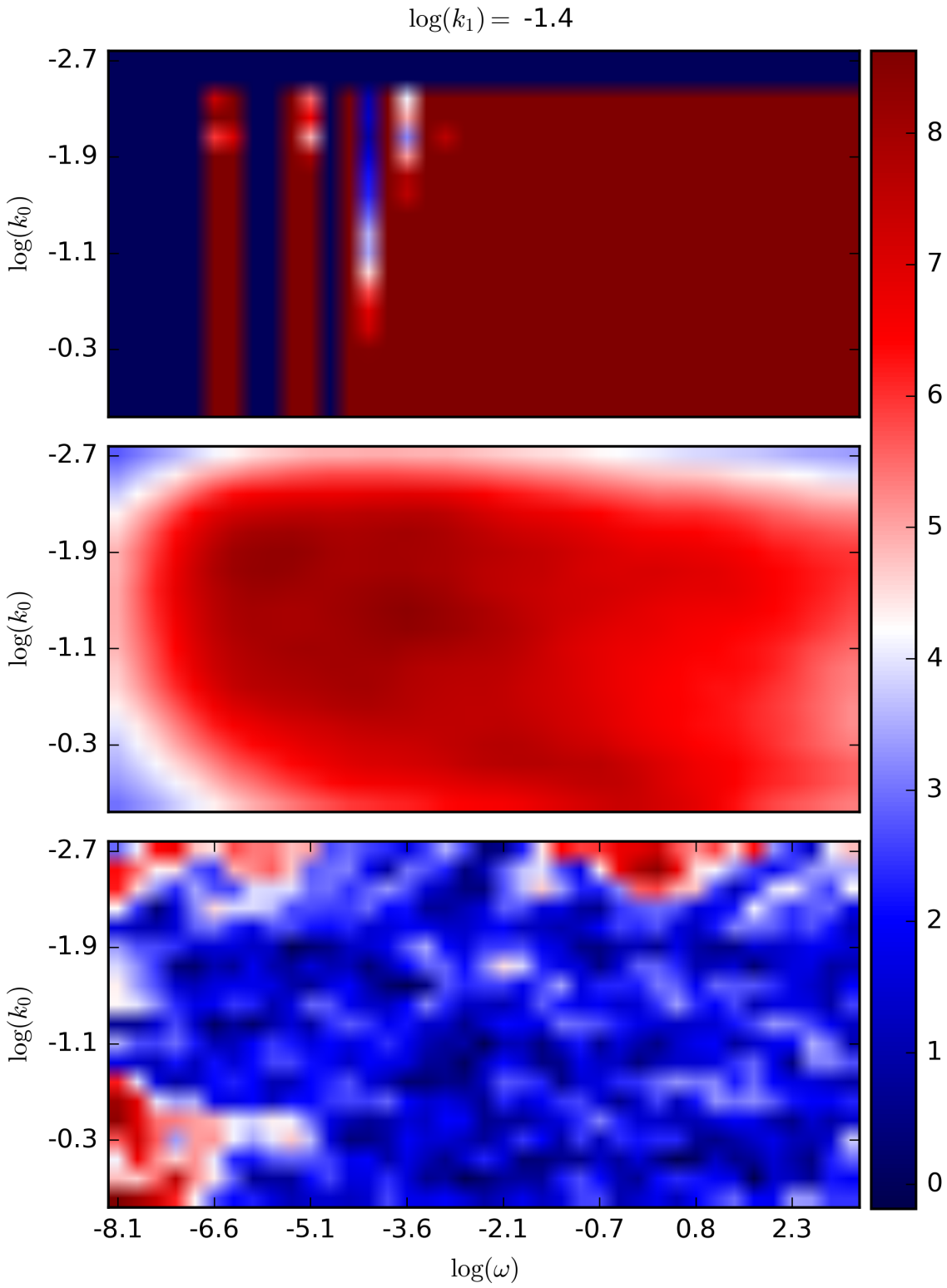

Figure 29: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the $\log(k_0)$-$\log(\omega)$-plane for a fixed $k1$. Compared to figure 21, spectral modes corresponding to space and time are a lot less correlated.

5.e Towards more sophisticated ionospheric heights

As we have already mentioned above, a major limitation to the reliability of the reconstructions is due to the simplified description of the ionosphere as a flat, two dimensional plane. In particular, the fact that we have to set a fixed height for this plane influences the reconstruction since the location of the projected data depends on this height. Due to the fact that we consider multiple station-source pairs simultaneously it is difficult to see in which way the reconstruction is affected by a change of the ionospheric height. Therefore, a more sophisticated solution would be to reconstruct this height together with all other fields of interest. However, this would again require to reformulate the data model which is beyond the scope of this work. To demonstrate the differences in the reconstruction arising from different heights, we also applied our reconstruction algorithm to the dataset d' using a height of 200 km. This height is closer to the reference height provided by the International Reference Ionosphere (IRI [28]) model, for the time period of the observation of this dataset. The resulting distortion fields are shown in figures 30 - 32 and the corresponding spectra in figures 33 - 35. Furthermore in figure 36 we depict the spatio-temporal evolution of the field and the corresponding spectrum in figure 37.

5.f Outlook

Aside from a reformulation of the data model in terms of a more realistic description of the ionosphere, there exist additional improvements to the reconstruction algorithm to overcome the preliminary nature of this application. One desired goal is to use the LOFAR data from all 125 different frequency channels. This reduces the effect of systematic errors due to the fact that there exists a known frequency dependence of the calibration phases which is of the form

$$\Delta\phi_R \propto \Delta\rho_e \;/\; \nu_R \;, \tag{43}$$

where $\Delta\phi_R$ and ν_R are the phase distortions and the frequency of the radio signal, and $\Delta\rho_e$ are fluctuations in the electron density of the ionosphere. Making use of this relation in a reformulated model, reduces systematic effects due to the fact that such systematics are not expected follow this relation.

In addition, the LOFAR stations provide another factor of 10 more in temporal resolution of the data, which, together with the data from all frequency channels, provides a large improvement in terms of the quality of the data and reduces the overall impact of measurement uncertainty.

The ultimate vision of this project is to fuse radio calibration with the ionospheric reconstructions provided by this work. This can improve the measurement of cosmic radio signals detected by LOFAR and could also be applied to upcoming radio interferometers such as the Square Kilometer Array (SKA).

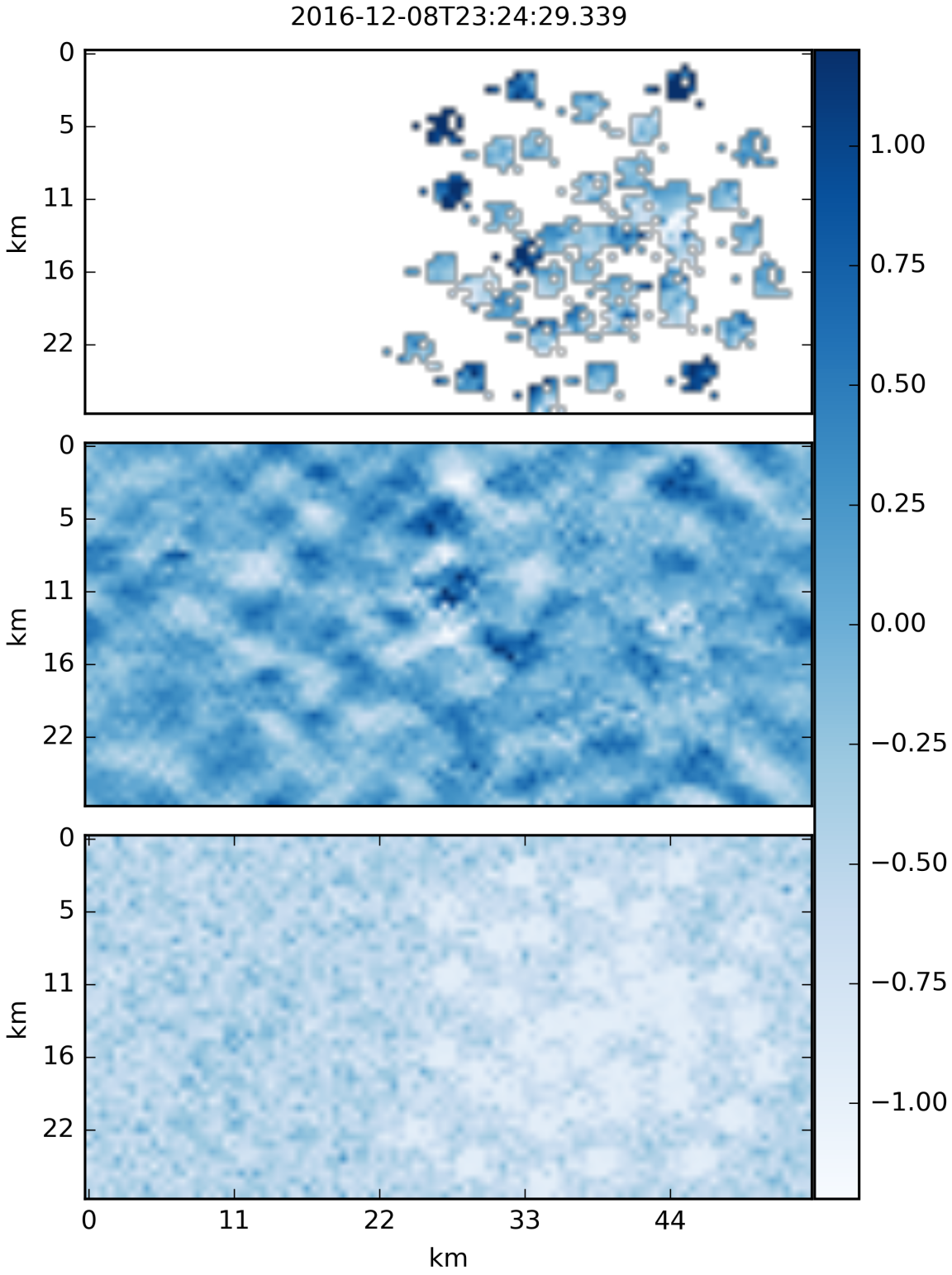

Figure 30: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the ionospheric plane, now located 200 km above the surface of the Earth. Again, the reconstruction was performed on the processed dataset d' (Eq. 42).

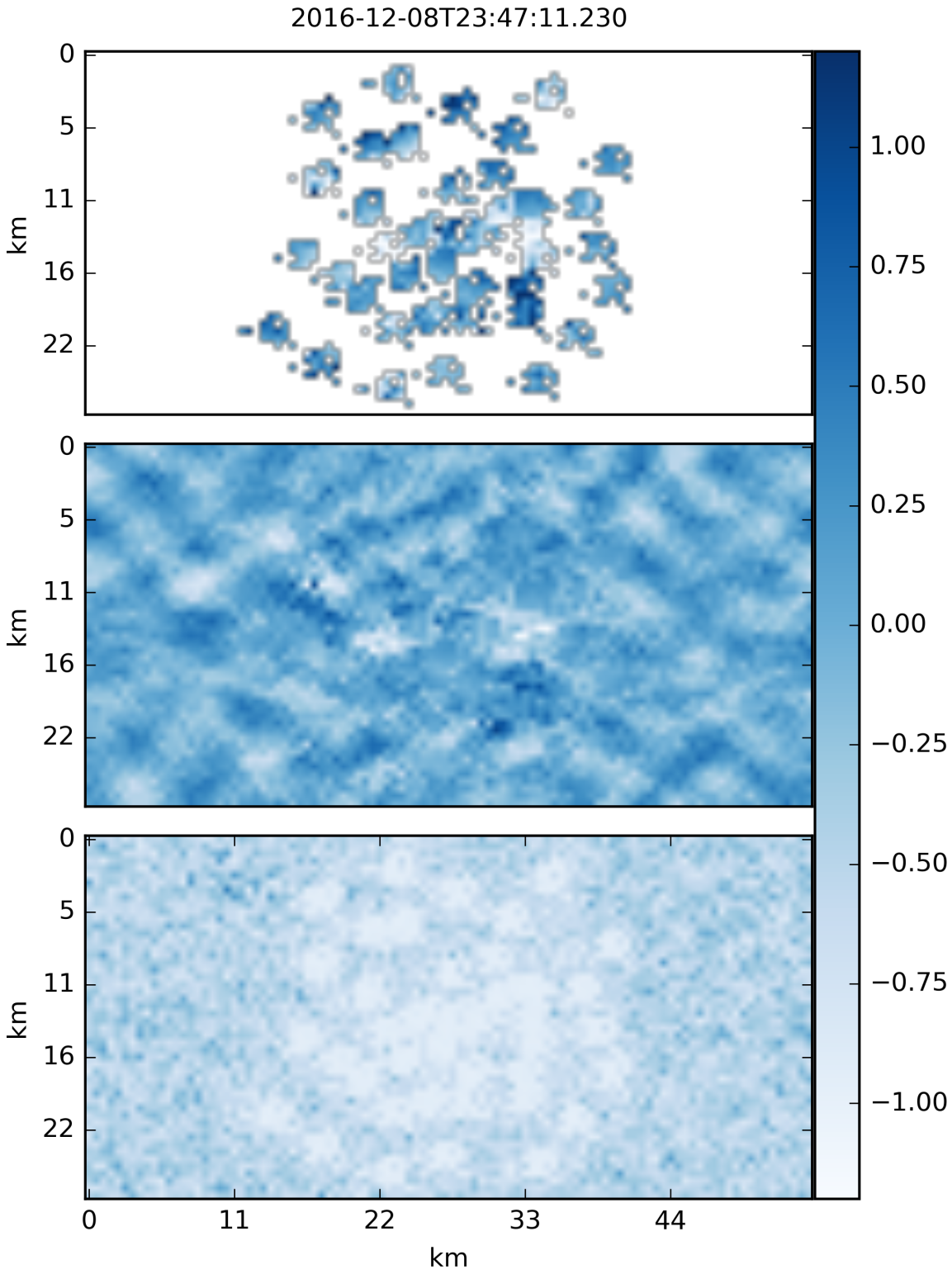

Figure 31: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the shifted ionospheric plane for a later moment in time compared to the previous figure.

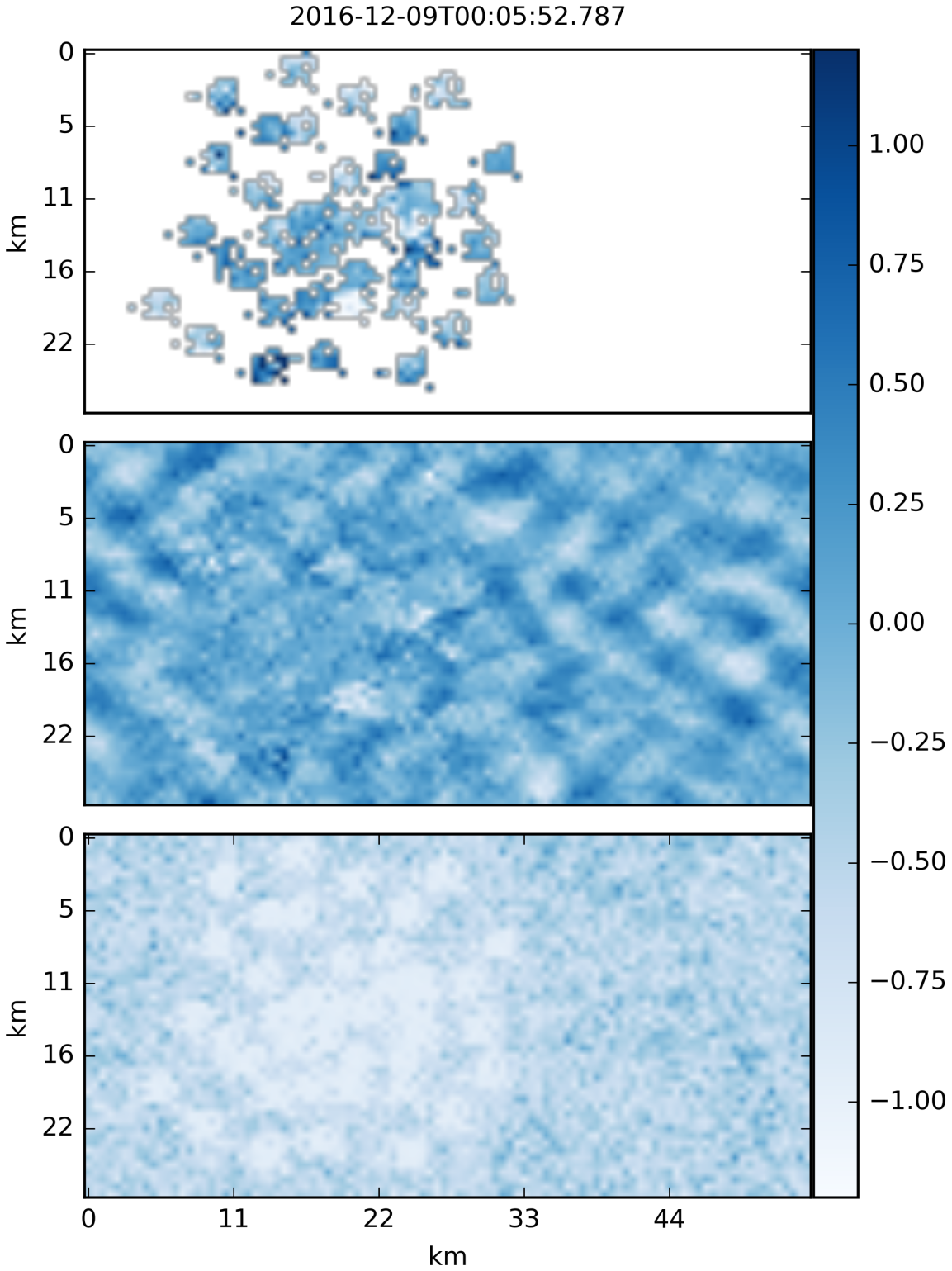

Figure 32: Data (top), MAP estimate (middle) and one sigma uncertainty (bottom) in the shifted ionospheric plane for a later moment in time compared to the previous figure.

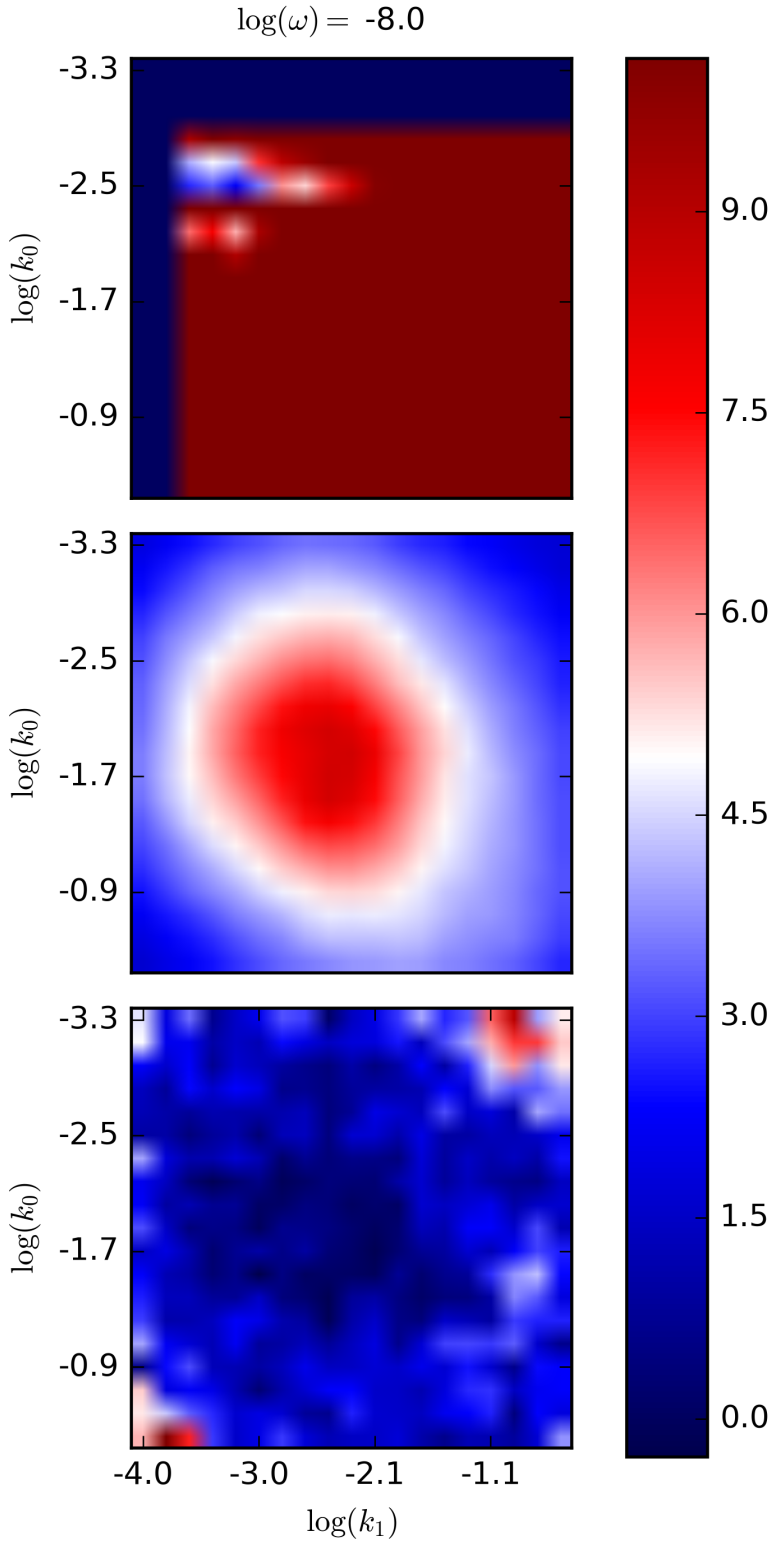

Figure 33: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for the setup used in the previous figures where we shifted the ionospheric plane to be located $200km$ above the surface of the Earth and used the processed dataset d'.

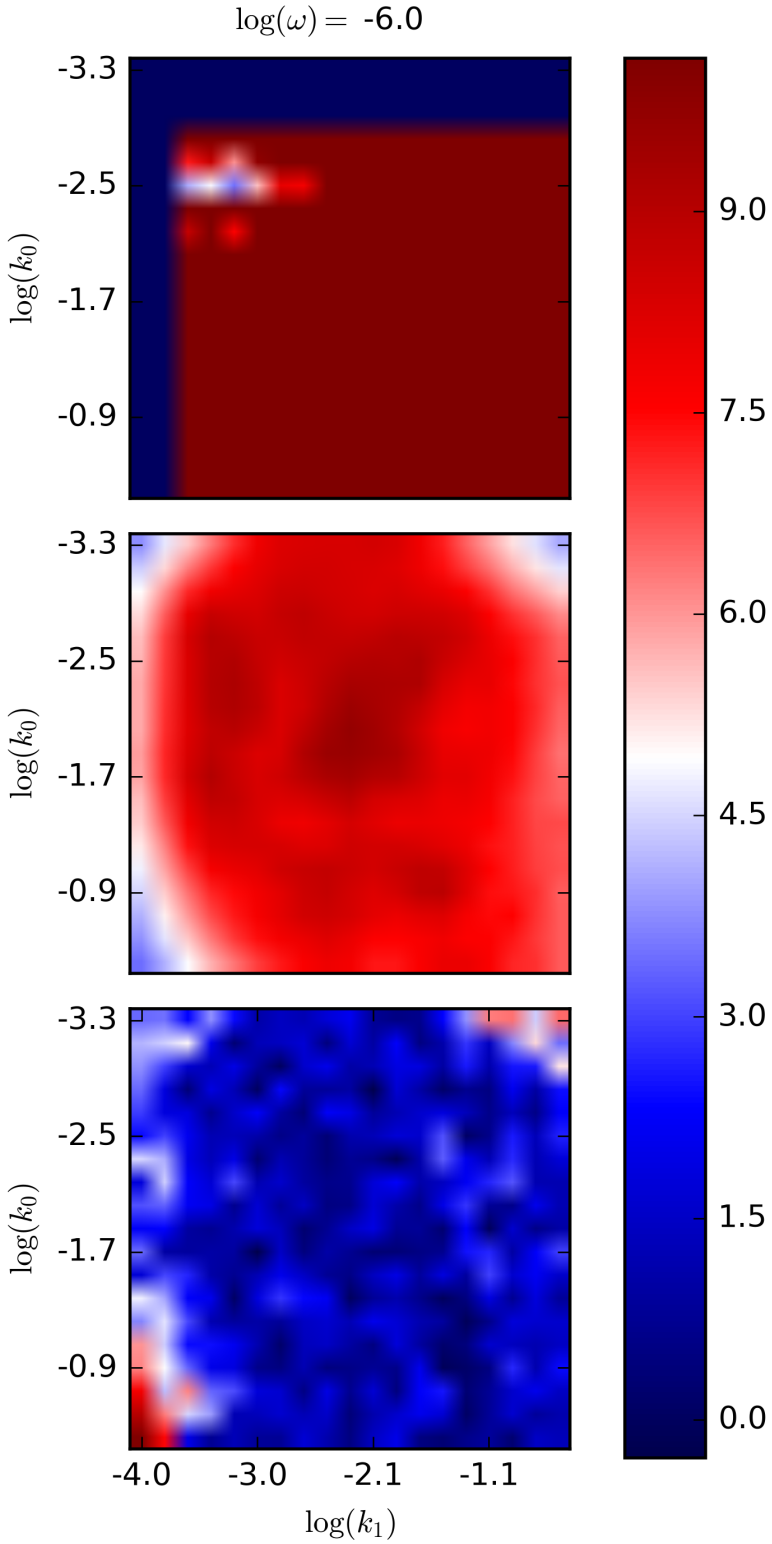

Figure 34: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for a larger value of ω compared to the previous figure.

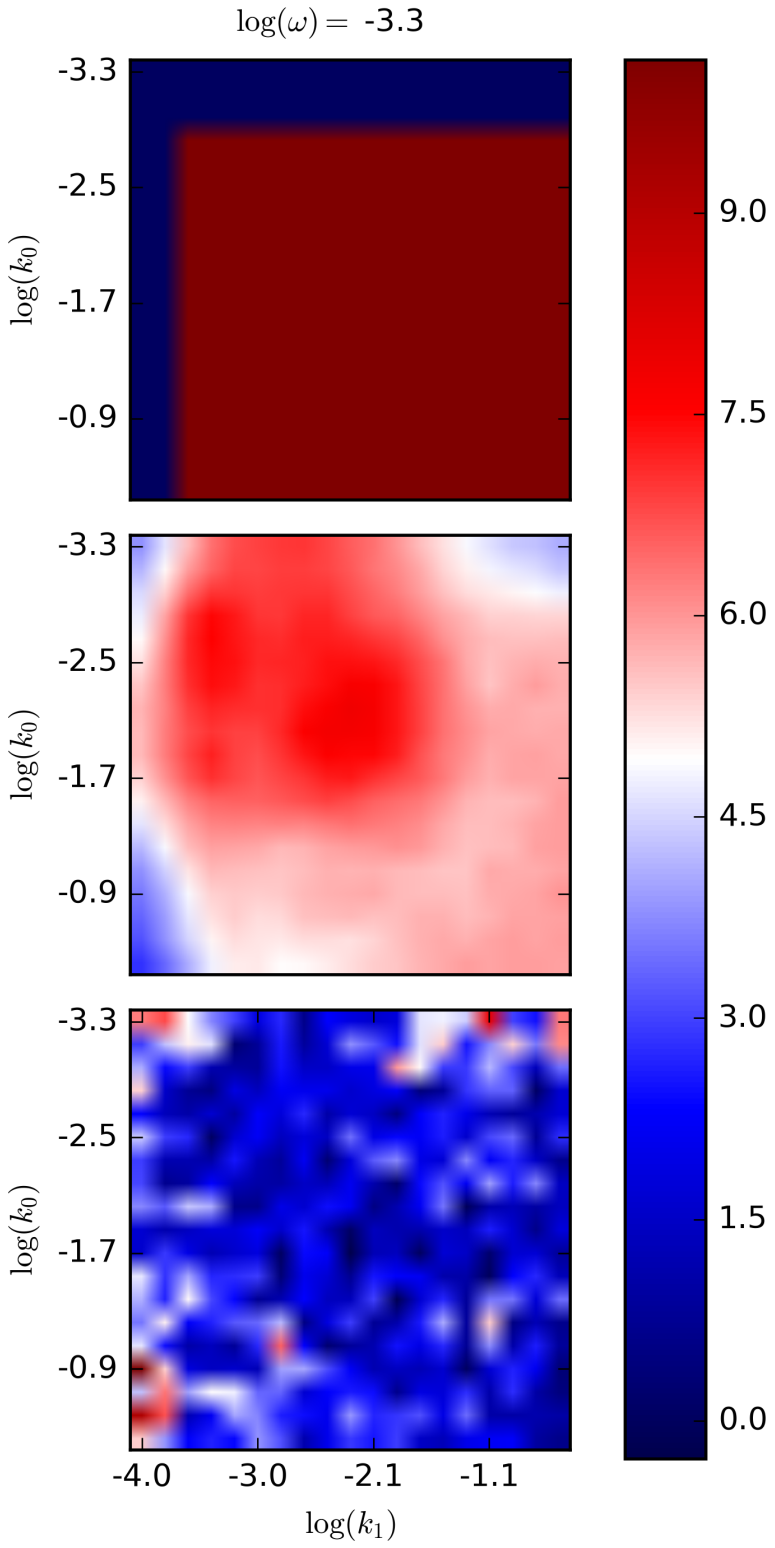

Figure 35: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the harmonic partner domain of the signal space on a log-log-scale for a larger value of ω compared to the previous figure.

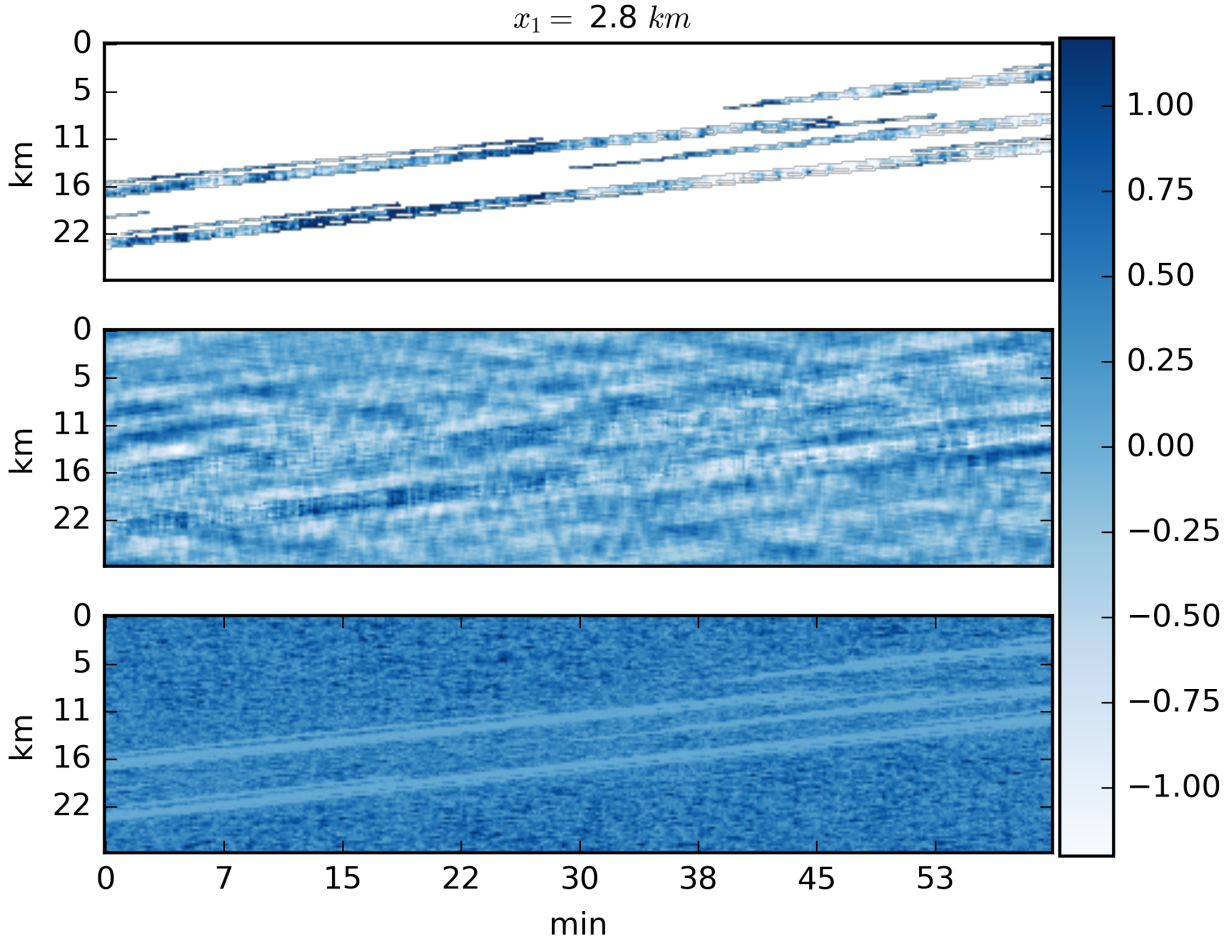

Figure 36: 1+1 dimensional spatio-temporal evolution of the data (top), the reconstruction (middle) and the corresponding one sigma uncertainty (bottom) for a fixed value of x_1, the other spatial coordinate. The ionospheric plane is located $200km$ above the surface of the Earth.

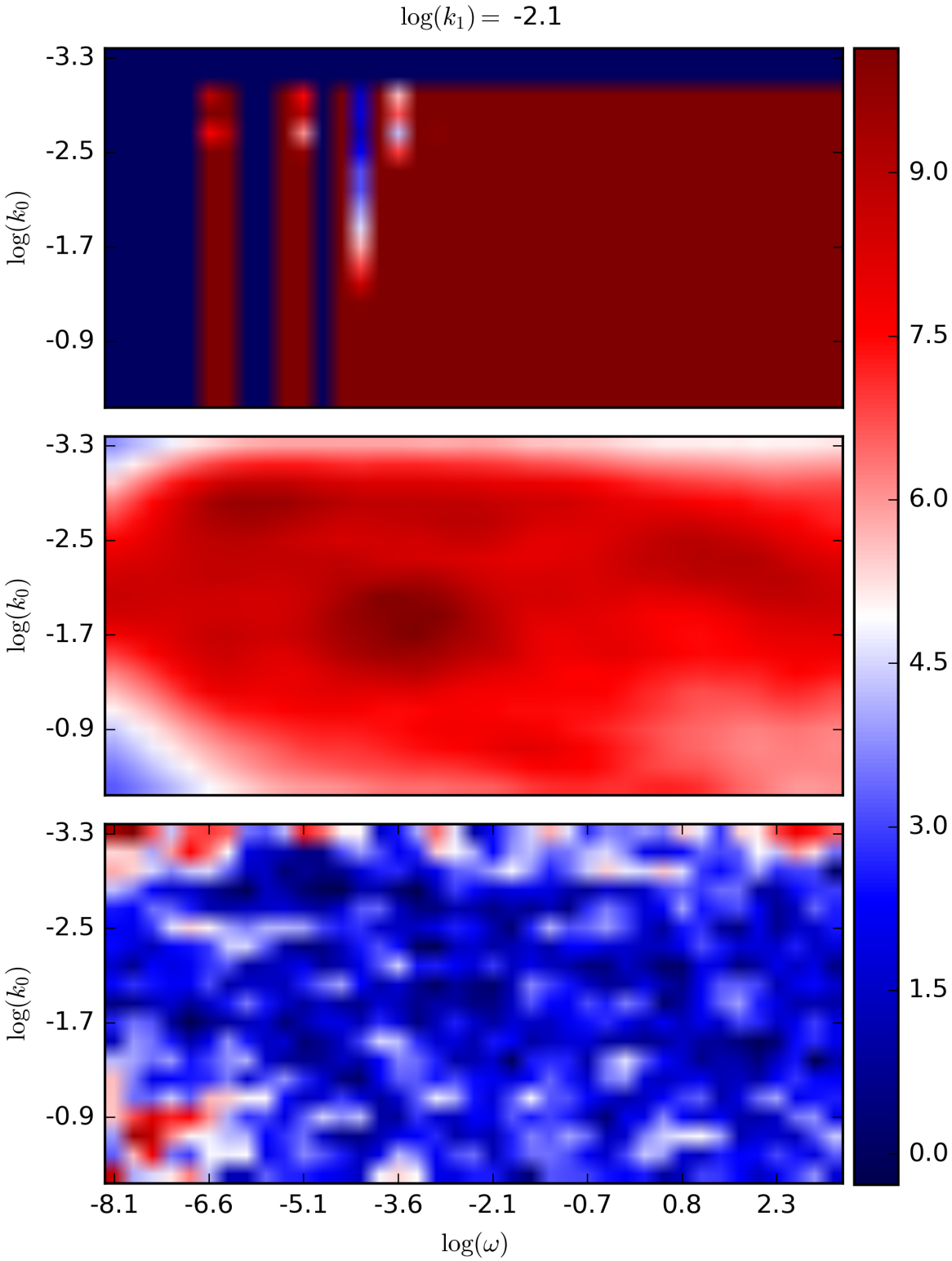

Figure 37: Projected data (top), spectral reconstruction (middle) and one sigma uncertainty (bottom) in the $\log(k_0)$-$\log(\omega)$-plane for a fixed $k1$.

6 Conclusion

Spectral density estimation is a powerful technique for field dynamics inference. In this work, posterior distributions for the spectral density and the field itself have been derived which ultimately are used to retrieve maximum a posteriori estimates as well as corresponding uncertainties. The one- and two-dimensional tests indicate that the method behaves as expected in self-consistent scenarios.

In addition, we demonstrated its applicability using calibration data of the LOFAR radio interferometers in order to reconstruct fluctuations in the ionosphere. Our method reconstructs these distortions in the ionospheric plane as a function of time, together with the spectral density of the underlying process. Note that this first application uses a simplified data model which renders all results to be preliminary. A more sophisticated model has to be used as there appear to be remaining systematic effects in the data, which currently still might dominate the reconstruction. However, a reformulation of this model only affects the way of how the data is being processed and consequently does not affect the applicability of the proposed method for multi-dimensional spectral density estimation.

Further possible applications of this method involve fields which have a non-trivial entanglement between spatial and temporal evolution. One example is the inference of the dynamics of a plasma from observations. Another example is the area of numerical simulations. In particular in astrophysical applications one is often interested in the dynamics of fields that also evolve on small scales, which are computationally too expensive to simulate. Treating a few expensive simulations as observations of the field of interest, our method can provide an approximate dynamics, an emulator that can mimic essential properties of the real evolution of the field.

The distinction between smooth and divergent parts of the spectrum as well as the corresponding properties of the prior choices for responsible fields τ and δ appears to be reasonable in this setting. However, other choices may also be possible. A future goal would be to study other parameterizations of the spectrum in terms of fields in particular in a 4D high resolution setting where a reconstruction of the full spectrum may exceed the range of computability.

Despite the fact that linear autonomous SDEs are an important class of SDEs to study dynamical evolution, an extension to non-autonomous as well as to non-linear problems is a desired goal for future work. However, for these cases it appears to be indispensable to have a general method for linear processes first. This is provided by this work.

7 References

[1] P. Frank, T. Steininger, and T. A. Enßlin. Field dynamics inference via spectral density estimation. *ArXiv e-prints*, 96(5):052104, November 2017.

[2] Douglas Henderson. *Stochastic Differential Equations in Science and Engineering*. World Scientific Publishing Company, 2006.

[3] Stefano M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples (Springer Series in Statistics)*. Springer Publishing Company, Incorporated, 1 edition, 2008.

[4] Stefan Klus, Nüske Peter, Koltai Peter, Wu Hao, Kevrekidis Ioannis, Schütte Christof, and Noé Frank. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, 2017. under review.

[5] Sam Roweis and Zoubin Ghahramani. A unifying review of linear gaussian models. *Neural Computation*, 11(2):305–345, February 1999.

[6] Yoann Potiron and Per Mykland. Local Parametric Estimation in High Frequency Data. Papers 1603.05700, arXiv.org, March 2016.

[7] Bernhard Spangl and Rudolf Dutter. *Robustness in Time Series: Robust Frequency Domain Analysis*, pages 207–223. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[8] Andreas Ruttor, Philipp Batz, and Manfred Opper. Approximate gaussian process inference for the drift function in stochastic differential equations. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2040–2048, 2013.

[9] G.L. Bretthorst. *Bayesian spectrum analysis and parameter estimation*. Lecture notes in statistics. Springer-Verlag, 1988.

[10] Christian Macaro and Raquel Prado. Spectral decompositions of multiple time series: A bayesian non-parametric approach. *Psychometrika*, 79(1):105–129, Jan 2014.

[11] Konstantin Knauf, Daniel Memmert, and Ulf Brefeld. Spatio-temporal convolution kernels. *Machine Learning*, 102(2):247–273, Feb 2016.

[12] T. Subba Rao and Gyorgy Terdik. A new covariance function and spatio-temporal prediction (kriging) for a stationary spatio-temporal random process. *Journal of Time Series Analysis*, 38(6):936–959, 2017. 10.1111/jtsa.12245.

[13] Yanbing Zheng, Jun Zhu, and Anindya Roy. Nonparametric bayesian inference for the spectral density function of a random field. *Biometrika*, 97(1):238–245, 2010.

[14] T. Enßlin. Information field theory. In U. von Toussaint, editor, *American Institute of Physics Conference Series*, volume 1553 of *American Institute of Physics Conference Series*, pages 184–191, August 2013.

[15] T. A. Enßlin, M. Frommert, and F. S. Kitaura. Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. *ArXiv e-prints*, 80(10):105005, November 2009.

[16] T. Steininger, J. Dixit, P. Frank, M. Greiner, S. Hutschenreuter, J. Knollmüller, R. Leike, N. Porqueres, D. Pumpe, M. Reinecke, M. Šraml, C. Varady, and T. Enßlin. NIFTy 3 - Numerical Information Field Theory - A Python framework for multicomponent signal inference on HPC clusters. *ArXiv e-prints*, August 2017.

[17] T. Enßlin. Astrophysical data analysis with information field theory. *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 1636:49–54, December 2014.

[18] J. Knollmüller and T. A. Enßlin. Noisy independent component analysis of autocorrelated components. *ArXiv e-prints*, May 2017.

[19] T. A. Enßlin. Information field dynamics for simulation scheme construction. *ArXiv e-prints*, 87(1):013308, January 2013.

[20] D. Pumpe, M. Greiner, E. Müller, and T. A. Enßlin. Dynamic system classifier. *ArXiv e-prints*, 94(1):012132, July 2016.

[21] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series, with engineering applications.* Number ix, 163 p. in Stationary time series. Technology Press of the Massachusetts Institute of Technology, 1950.

[22] M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows, M. Brentjens, A. Nelles, R. Beck, H. Falcke, R. Fender, J. Hörandel, L. V. E. Koopmans, G. Mann, G. Miley, H. Röttgering, B. W. Stappers, R. A. M. J. Wijers, S. Zaroubi, M. van den Akker, A. Alexov, J. Anderson, K. Anderson, A. van Ardenne, M. Arts, A. Asgekar, I. M. Avruch, F. Batejat, L. Bähren, M. E. Bell, M. R. Bell, I. van Bemmel, P. Bennema, M. J. Bentum, G. Bernardi, P. Best, L. Bîrzan, A. Bonafede, A.-J. Boonstra, R. Braun, J. Bregman, F. Breitling, R. H. van de Brink, J. Broderick, P. C. Broekema, W. N. Brouw, M. Brüggen, H. R. Butcher, W. van Cappellen, B. Ciardi, T. Coenen, J. Conway, A. Coolen, A. Corstanje, S. Damstra, O. Davies, A. T. Deller, R.-J.

Dettmar, G. van Diepen, K. Dijkstra, P. Donker, A. Doorduin, J. Dromer, M. Drost, A. van Duin, J. Eislöffel, J. van Enst, C. Ferrari, W. Frieswijk, H. Gankema, M. A. Garrett, F. de Gasperin, M. Gerbers, E. de Geus, J.-M. Grießmeier, T. Grit, P. Gruppen, J. P. Hamaker, T. Hassall, M. Hoeft, H. A. Holties, A. Horneffer, A. van der Horst, A. van Houwelingen, A. Huijgen, M. Iacobelli, H. Intema, N. Jackson, V. Jelic, A. de Jong, E. Juette, D. Kant, A. Karastergiou, A. Koers, H. Kollen, V. I. Kondratiev, E. Kooistra, Y. Koopman, A. Koster, M. Kuniyoshi, M. Kramer, G. Kuper, P. Lambropoulos, C. Law, J. van Leeuwen, J. Lemaitre, M. Loose, P. Maat, G. Macario, S. Markoff, J. Masters, R. A. McFadden, D. McKay-Bukowski, H. Meijering, H. Meulman, M. Mevius, E. Middelberg, R. Millenaar, J. C. A. Miller-Jones, R. N. Mohan, J. D. Mol, J. Morawietz, R. Morganti, D. D. Mulcahy, E. Mulder, H. Munk, L. Nieuwenhuis, R. van Nieuwpoort, J. E. Noordam, M. Norden, A. Noutsos, A. R. Offringa, H. Olofsson, A. Omar, E. Orrú, R. Overeem, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, J.-P. de Reijer, J. Reitsma, G. A. Renting, P. Riemers, E. Rol, J. W. Romein, J. Roosjen, M. Ruiter, A. Scaife, K. van der Schaaf, B. Scheers, P. Schellart, A. Schoenmakers, G. Schoonderbeek, M. Serylak, A. Shulevski, J. Sluman, O. Smirnov, C. Sobey, H. Spreeuw, M. Steinmetz, C. G. M. Sterks, H.-J. Stiepel, K. Stuurwold, M. Tagger, Y. Tang, C. Tasse, I. Thomas, S. Thoudam, M. C. Toribio, B. van der Tol, O. Usov, M. van Veelen, A.-J. van der Veen, S. ter Veen, J. P. W. Verbiest, R. Vermeulen, N. Vermaas, C. Vocks, C. Vogt, M. de Vos, E. van der Wal, R. van Weeren, H. Weggemans, P. Weltevrede, S. White, S. J. Wijnholds, T. Wilhelmsson, O. Wucknitz, S. Yatawatta, P. Zarka, A. Zensus, and J. van Zwieten. LOFAR: The LOw-Frequency ARray. *ArXiv e-prints*, 556:A2, August 2013.

[23] N. Oppermann, M. Selig, M. R. Bell, and T. A. Enßlin. Reconstruction of Gaussian and log-normal fields with spectral smoothness. *ArXiv e-prints*, 87(3):032136, March 2013.

[24] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995.

[25] J. N. Bahcall and R. K. Ulrich. Solar models, neutrino experiments, and helioseismology. *Reviews of Modern Physics*, 60:297–372, April 1988.

[26] K. Davies and Institution of Electrical Engineers. *Ionospheric Radio*. Electromagnetics and Radar Series. Peregrinus, 1990.

[27] M.C. Kelley. *The Earth's Ionosphere: Plasma Physics and Electrodynamics*. International Geophysics. Elsevier Science, 2009.

[28] K Rawer, Dieter Bilitza, and S Ramakrishnan. *Goals and status of the International Reference Ionosphere*, volume 16. 05 1978.

[29] G.E. Forsythe and W.R. Wasow. *Finite-difference methods for partial differential equations*. Applied mathematics series. Wiley, 1960.

A Smoothness Prior in higher dimensions

A smoothness prior in higher dimensions is sometimes constructed by imposing the constraint that, at each point y, the log-Laplacian $\partial^2/\partial(\log(y))^2$ of the field should be small. This however, is not sufficient to impose smoothness in cases where one encounters concave and convex curvature along different directions simultaneously (e.g., saddle surfaces). In the following we briefly outline why this is the case.

If we consider the logarithmic Hessian of a field ψ at each point $\mathbf{y} \in \mathbb{R}^N$, defined as

$$(H[\psi])_{ij}(\mathbf{y}) = \frac{\partial^2 \psi(\mathbf{y})}{\partial \log(y_i)\ \partial \log(y_j)}\ ,\ i,j \in \{1,...,N\}\ , \tag{44}$$

we notice that the log-Laplacian is equal to the trace of the Hessian and therefore to the sum of the corresponding eigenvalues. This indicates that if the eigenvalues are positive and negative (corresponding to convex and concave curvature along the eigendirections), the Laplacian can become zero even though the surface has nonzero curvature. As a result, saddle-surfaces with equal absolute curvature and constant surfaces are equally likely in the corresponding prior. This is not a desired behavior. The goal of a generic smoothness-prior should be to assign lower probabilities also to surfaces with altering curvature.

We therefore propose to use a prior which aims to minimize all quadratic, logarithmic variations of a field ψ simultaneously. For a M-dimensional space the exponential factor of the prior reads

$$\psi^\dagger T_\sigma^{-1} \psi = \frac{1}{\sigma^2}\ \psi^\dagger \sum_{i=1}^{M} \left(T_{ii}^{-1} + 2 \sum_{j=1}^{i-1} T_{ij}^{-1} \right) \psi\ , \tag{45}$$

with T_{ij} such that

$$\psi^\dagger T_{ij}^{-1} \psi = \int d^M \left(\log\left(\mathbf{y}\right)\right)\ \left|(H[\psi])_{ij}(\mathbf{y})\right|^2\ . \tag{46}$$

Note that this prior is rotationally invariant which indicates that curvature in all directions is treated in the same way.

A.a Discrete derivatives

In order to apply the theoretic discussions above to a finite setting (e.g., a finite grid on a computer) we need a discrete representation of the operators involved in the analysis,

in particular of the log-derivative operator. One usual way is to approximate derivatives in terms of finite differences (see e.g., [29]). A possible way of discretizing the second logarithmic derivative is described in the appendix of [23].

However, in this particular setting there exist also another way of differentiation in terms of Fourier transformation. We note that, using the chain rule, the second logarithmic derivative of y can be written as

$$\partial^2_{\log(y)} = y^2 \partial^2_y + y \partial_y \ , \tag{47}$$

where we restrict our discussion to one dimension, for simplicity. Furthermore, as discussed in section 2, differentiation can be translated to multiplication in harmonic space. I.e.

$$\partial_y \psi(y) = \int dk \ ik \ \tilde{\psi}(k) \ e^{iky} \ , \tag{48}$$

which indicates that using discrete Fourier transformations, derivatives can be represented by point-wise multiplication in harmonic space.

At first sight, this may seem like a more complicated method for differentiation. However, we note that on a parallel machine, Fourier transformations as well as point-wise multiplications can be fully parallelized while finite differences methods always involve inter-node communication due to subtracting shifted versions of the field representation. On a single node, however, finite differences appear to be computationally more efficient.

Therefore, depending on the problem setting, one method may be superior to the other. Our tests indicate that both methods of differentiation are applicable for prior construction in our problem setting. However, we do not recommend to mix both methods within one inference problem, as the exact form of the derivatives might be incompatible. A more sophisticated test in terms of computational time and accuracy is beyond the scope of this work.

B Complex logarithm and smoothness at zero

In some applications of the spectral density inference method, we need to impose the smoothness-prior on a zero-centered harmonic space, since the negative part of the density can carry additional information and a shift to purely positive values is not always possible.

As the smoothness-prior involves logarithmic derivatives, we seek to find a way to define logarithmic derivatives for negative values. This is achieved in terms of the complex logarithm. Consider for example $k > 0$, then

$$\log(-k) = \log(e^{i\pi} k) = \log(k) + i\pi \ , \tag{49}$$

and therefore the infinitesimal line-element reads

$$|d\log(-k)| = |d\log(k)| \ , \ \forall k \neq 0 \ . \tag{50}$$

This indicates that we can express the derivative with respect to a negative k in terms of the corresponding positive differential, i.e.

$$\left|\frac{\partial\psi(k)}{\partial\log(k)}\right| = \left|\frac{\partial\psi(k)}{\partial\log(|k|)}\right| \ , \ \forall k \neq 0 \ . \tag{51}$$

As the logarithm of zero is not defined, the smoothness-prior is also not defined at zero. In this work, we fix this problem by adding a prior to the analysis which aims to minimize the second derivative w.r.t k. Therefore, the Hamiltonians of τ and δ get modified by a term

$$H_\eta(\psi) = \frac{1}{2}\psi^\dagger D_\eta^{-1}\psi = \frac{1}{2\eta^2}\int\left|\frac{\partial^2\psi(k)}{\partial k^2}\right|^2 dk \ , \ \psi \in \{\tau,\delta\} \ . \tag{52}$$

Note that for small k this prior dominates the smoothness prior, while for larger k the logarithmic derivatives are dominant and this second prior does not contribute significantly any more. All applications shown in this work use $\eta = 0.1$.

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

Ort, Datum: München, 13.03.2018 Unterschrift: