Nonlin. Processes Geophys., 16, 197–210, 2009
www.nonlin-processes-geophys.net/16/197/2009/

Nonlinear Processes in Geophysics

Image-model coupling: a simple information theoretic perspective for image sequences

N. D. Smith[1,*], C. N. Mitchell[2], and C. J. Budd[1]

[1]Department of Mathematical Sciences, University of Bath, BA2 7AY, UK
[2]Department of Electronic and Electrical Engineering, University of Bath, BA2 7AY, UK
[*]Completed while with the Department of Electronic and Electrical Engineering, University of Bath, UK

Abstract. Images are widely used to visualise physical processes. Models may be developed which attempt to replicate those processes and their effects. The technique of coupling model output to images, which is here called "image-model coupling", may be used to help understand the underlying physical processes, and better understand the limitations of the models. An information theoretic framework is presented for image-model coupling in the context of communication along a discrete channel. The physical process may be regarded as a transmitter of images and the model as part of a receiver which decodes or recognises those images. Image-model coupling may therefore be interpreted as image recognition. Of interest are physical processes which exhibit "memory". The response of such a system is not only dependent on the current values of driver variables, but also on the recent history of drivers and/or system description. Examples of such systems in geophysics include the ionosphere and Earth's climate. The discrete channel model is used to help derive expressions for matching images and model output, and help analyse the coupling.

1 Introduction

Images describe what is present in a scene, and mathematical or empirical models may be developed which attempt to replicate the underlying processes which produce those images. Given an image, it is interesting to find the model input producing output which best matches the image. In this way useful knowledge concerning the mechanisms producing that image may be discerned. Such image-model coupling, in the context of sequences of images and systems with memory, is described below. The particular example application chosen is the coupling of images of the ionosphere and the output of ionospheric models. This is a challenging problem because the ionosphere (Hargreaves, 2003) is a nonlinear system with regard to the response of its electron content to its driver variables. Furthermore, the system has "memory" since its response at a given time depends not only on the current drivers but also on the recent history.

There are various questions to consider. What objective function should be used to describe the match between an image sequence and model output? More generally, how can competing models be compared in a quantitative manner? And how can the importance of driver variables, for replicating image sequences, be assessed?

To attempt to answer these questions in a consistent manner, image-model coupling is presented within a simple information theoretic context as transmission along a discrete channel. The simpler and more familiar discrete memoryless channel is a special case suitable for systems which do not exhibit memory. This approach clearly separates the true underlying real-world process producing the image sequence and the proposed model, and gives a framework to help identify assumptions in the proposed model and possible sources

Correspondence to: N. D. Smith
(n.smith@bath.ac.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union and the American Geophysical Union.

of error in coupling model output to images. The objective function used for coupling is statistical in nature; with an increased availability of data, it should be possible to derive objective functions which improve the coupling. The context encourages the use of "tools" drawn from information theory and statistical modelling.

The approach borrows heavily from joint source-channel decoding (JSCD), e.g. Garcia-Frias and Villasenor (2001); Ferrari et al. (2005); Link and Kallel (2000) and speech recognition, e.g. Rabiner and Juang (1993), where both maximum a-posteriori (MAP) decoding and evaluation in terms of error rate are commonplace. Modelling channel noise is important for communication systems in general, e.g. Burlina and Alajaji (1998); Beaulieu (1991). Our approach is very similar to the analysis of stochastic channels in Ferrari et al. (2005), except we have here assumed specific conditional independences. This paper seeks to apply concepts from information theory and speech recognition to the modelling of geophysical systems; memory is often important for such systems. The ionosphere (e.g. see Hargreaves, 2003) is used as an illustration in this text. Memory is also important for Earth's climate, e.g. Maraun et al. (2004). The statistical aspects of our framework are embedded in graphical models and the generic Bayesian approach; regarding geophysical systems, note that graphical models, hierarchical Bayesian models and hidden Markov models (HMMs) have previously been applied, e.g. Ihler et al. (2007); Wikle et al. (1998); Beyreuther and Wassermann (2008). There is significant overlap with data assimilation, e.g. Wikle and Berliner (2007); however data assimilation is primarily interested in using models to improve the prediction of observations rather than to decode the values of driver variables. Ignoring this distinction, image-model coupling is essentially a form of data assimilation. Our distinct separation of real-world process and proposed model is similar to that in Duane et al. (2006), where data assimilation is viewed in terms of the synchronisation of two dynamical systems via a noisy communication channel. There are similarities between the framework for data assimilation applied to the ionosphere in Guo et al. (2003) and our approach, but in Guo et al. (2003) there is no development in terms of a discrete communication channel. As far as the authors are aware, the main contribution of this paper is in the application of concepts from JSCD, speech recognition and statistical modelling to the "decoding" of the values of driver variables for geophysical systems, in particular by highlighting the role of memory and missing variables, and in analysing the coupling. While the concepts are not new, the authors hope the approach may encourage extra insight in modelling geophysical systems.

The paper is organised as follows. First, Sect. 2 introduces discrete channel models with and without memory; the true underlying process producing the image sequence is considered as a transmitter, and the proposed model is part of a receiver which interprets, or decodes, the image sequence. Section 3 then describes the receiver in more detail; the codebook, noise and state transition models, objective function for matching, and the search mechanism. In particular, the assumptions implicit in applying simple sum square error minimisation are detailed. Sections 4 and 5 respectively introduce expressions to compare alternative models, and estimate the sensitivity of image sequences to different driver variables. Finally some discussion and conclusions follow in Sects. 6 and 7.

2 Communication channel framework

Image-model coupling may be viewed from an information theoretic perspective as communication via a discrete channel (e.g. see MacKay, 2004; Abramson, 1963; Khinchin, 1957). The true-world process generating the images is viewed as a transmitter, and the model which is used to interpret the images is part of the receiver. Some form of synchronous "online" decoding is attractive for continuous communication, i.e. continuous "online" image-model coupling. In the context of ionospheric modelling, the true ionosphere acts as a transmitter and encodes physical driver variables as an image of the ionosphere, where the image is simply an often incomplete description of the ionosphere. The image may be in the form of electron densities or their line integrals. The ionospheric model is part of a decoder which attempts to recover the values of those driver variables. This channel approach allows us to derive an objective function for matching a temporal sequence of images with model output. The following analysis concerns systems such as the ionosphere which are not memoryless, so receivers which assume simple discrete memoryless channel models may not be accurate. However, for reasons of tractability, such receivers may be applied, though it is useful to understand the limitations and assumptions in so doing. In the following, "TX" denotes the transmitter and "RXq" the receiver for $q \in \{3, 2, 1\}$.

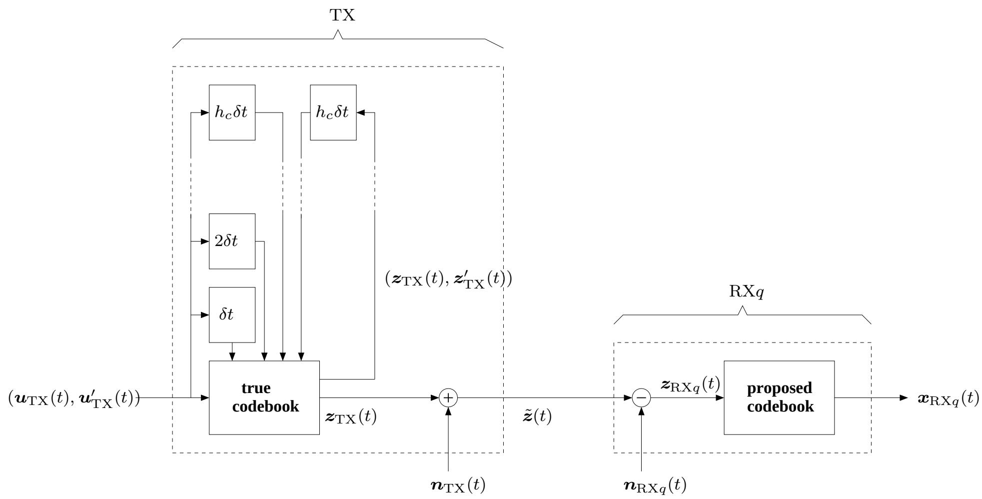

Fig. 1. Image-model coupling as a discrete channel model, where the true real-world process is an encoder, δt is the duration of a timestep when used to describe delay in a temporal buffer, and RXq, $q\in\{3, 2, 1\}$ denotes different receivers.

2.1 Transmitter (TX)

Figure 1 describes the true real-world process assumed to generate the images. The process is assumed driven. The driver variables of interest are recorded at time t as $\boldsymbol{u}_{\rm TX}(t)\in U_{\rm TX}$, where $U_{\rm TX}$ is a discrete, typically open set[1]. For the ionosphere, driver variables of interest may include measurements of solar or geomagnetic activity. In addition, there are latent driver variables $\boldsymbol{u}'_{\rm TX}(t)\in U'_{\rm TX}$ which are not measured and are typically unknown or not of direct interest to the modeller. Again, $U'_{\rm TX}$ is a discrete set. The current system is fully described by $z_{\rm TX}(t)\in Z_{\rm TX}$ and $z'_{\rm TX}(t)\in Z'_{\rm TX}$, which respectively describe those variables which form the image, and the complementary set of variables required to complete the full description. For convenience, define discrete sets $Z_{\rm TX}\subseteq Z$ and $Z'_{\rm TX}\subseteq Z'$. The real-world process is assumed viewed as a codebook which implements the deterministic mapping, for some fixed and known channel memory length[2] $h_c\in\mathbb{N}_0$ (for channel memory, e.g. see Khinchin, 1957),

$$z_{\rm TX}(t) : (\boldsymbol{u}_{\rm TX}(t-h_c, t), \boldsymbol{u}'_{\rm TX}(t-h_c, t), z_{\rm TX}(t-h_c), z'_{\rm TX}(t-h_c)) \mapsto z_{\rm TX}(t), \quad (1)$$

where t is the timestep index and h_c is expressed in timesteps, and for example,

$$\boldsymbol{u}_{\rm TX}(t-h_c, t)\equiv(\boldsymbol{u}_{\rm TX}(t-h_c), \boldsymbol{u}_{\rm TX}(t-h_c+1), \ldots, \boldsymbol{u}_{\rm TX}(t)).$$

Similar abbreviations are used elsewhere for time-ordered temporal sequences of vectors. Hence $z_{\rm TX}(t)$ is fully determined by the present and past driver variables, both known and latent, and an initial complete description. Hence there are no variables beyond those in the domain of the mapping which cause variation in $z_{\rm TX}(t)$. The assumption of the deterministic mapping is thought reasonable due to the inclusion of all driver variables and description variables in the domain. In the case of the ionosphere, the dependency on initialisation is assumed since the ionosphere is not memoryless; its description may evolve differently over time under the action of the same sequence of drivers depending on the initial distribution of plasma.

There is also memory in the source. Letting $h_s\in\mathbb{N}_0$ denote the length of source memory in timesteps, then it is convenient to define the state,

$$\boldsymbol{x}_{\rm TX}(t) = (\boldsymbol{u}_{\rm TX}(t-h, t), \boldsymbol{u}'_{\rm TX}(t-h, t), z_{\rm TX}(t-h), z'_{\rm TX}(t-h)), \quad (2)$$

where $h{=}\max[h_s{-}1, h_c]$, where $\boldsymbol{x}_{\rm TX}(t)\in X_{\rm TX}$, and where $X_{\rm TX}{=}(\otimes_{i=1}^{h+1} U_{\rm TX})(\otimes_{i=1}^{h+1} U'_{\rm TX}) \otimes Z_{\rm TX} \otimes Z'_{\rm TX}$. This definition of state allows the transmitter to be modelled as a hidden Markov process (see Sect. 3.4).

[1]The constraint of discrete signals and sets rather than continuous analogues is necessary for a discrete channel model; although real-world processes are typically continuous, discretisation may be regarded as the result of sampling continuous signals or spaces into the machine precision of the recording, storage or computing device.

[2]The notation $\mathbb{N}_0$ denotes all positive integers and the zero.

The actual image measured or recorded at time t is $\tilde{z}(t)\in Z$. This image is related to the partial description $z_{\mathrm{TX}}(t)$ by,

$$\tilde{z}(t) = z_{\mathrm{TX}}(t) + \boldsymbol{n}_{\mathrm{TX}}(t), \tag{3}$$

where $\boldsymbol{n}_{\mathrm{TX}}(t)\in Z$ is additive noise describing error in the measuring devices. Unfortunately, if an image is incomplete, it is sometimes necessary to complete the image, for example using tomographic reconstruction. For the purposes of this analysis, such images are regarded as if directly imaged by a device, and the error in the reconstruction included into the noise process $\boldsymbol{n}_{\mathrm{TX}}(t)$. Both the true real-world process and noise source may be nonstationary; however for the estimation of the statistical models described later, properties of stationarity and ergodicity (Korn and Korn, 1968) are convenient. The noise signal $\boldsymbol{n}_{\mathrm{TX}}(t)$ is not transmitted independently along the channel, only the image $\tilde{z}(t)$. In practice, it is usual to consider a finite length sequence of images, for example T images $\tilde{z}(1, T)$.

2.2 Level 3 receiver (RX3)

With infinite knowledge, it is possible to construct a receiver which implements the reverse process to the transmitter. Hence the receiver first "denoises" the image,

$$z_{\mathrm{RX3}}(t) = \tilde{z}(t) - \boldsymbol{n}_{\mathrm{RX3}}(t), \tag{4}$$

where $\boldsymbol{n}_{\mathrm{RX3}}(t)\in Z$, $z_{\mathrm{RX3}}(t)\in Z_{\mathrm{RX3}}$, and Z_{RX3} is typically a discrete set denoting the range of the receiver codebook. The codebook implements deterministic mappings of form,

$$z_{\mathrm{RX3}}(t) : (\boldsymbol{u}_{\mathrm{RX3}}(t-h_c, t), \boldsymbol{u}'_{\mathrm{RX3}}(t-h_c, t), z_{\mathrm{RX3}}(t-h_c), z'_{\mathrm{RX3}}(t-h_c)) \mapsto z_{\mathrm{RX3}}(t), \tag{5}$$

but is used in the reverse direction. The codebook is typically implemented by a deterministic mathematical or empirical model. The noise source $\boldsymbol{n}_{\mathrm{RX3}}(t)$ should model the measurement noise in the imaging devices. This receiver is unrealisable, but is included since it permits decoding with the lowest possible error rate. Decoding is described more fully in Sect. 3.4. For clarity, the receiver is here called a level 3 receiver, where the higher the level, the deeper the conditional dependencies in the receiver codebook. For convenience, a receiver state is defined,

$$\boldsymbol{x}_{\mathrm{RX3}}(t) = (\boldsymbol{u}_{\mathrm{RX3}}(t-h, t), \boldsymbol{u}'_{\mathrm{RX3}}(t-h, t), z_{\mathrm{RX3}}(t-h), z'_{\mathrm{RX3}}(t-h)), \tag{6}$$

where $h = \max[h_s - 1, h_c]$ and $\boldsymbol{x}_{\mathrm{RX3}}(t) \in X_{\mathrm{RX3}} = (\otimes_{i=1}^{h+1} U_{\mathrm{RX3}})(\otimes_{i=1}^{h+1} U'_{\mathrm{RX3}}) \otimes Z_{\mathrm{RX3}} \otimes Z'_{\mathrm{RX3}}$.

2.3 Level 2 receiver (RX2)

This is similar to the level 3 receiver with differences in the codebook mappings and notation. Denoising is,

$$z_{\mathrm{RX2}}(t) = \tilde{z}(t) - \boldsymbol{n}_{\mathrm{RX2}}(t), \tag{7}$$

where $\boldsymbol{n}_{\mathrm{RX2}}(t)\in Z$ and $z_{\mathrm{RX2}}(t)\in Z_{\mathrm{RX2}}$. The codebook implements deterministic mappings of form,

$$z_{\mathrm{RX2}}(t) : (\boldsymbol{u}_{\mathrm{RX2}}(t-h_c, t), z_{\mathrm{RX2}}(t-h_c, t-1)) \mapsto z_{\mathrm{RX2}}(t), \tag{8}$$

where the unmeasured or unknown driver and description variables in U'_{TX} and Z'_{TX} have been omitted. The stochastic variation in $\tilde{z}(t)$ due to these variables is instead incorporated into a more complicated noise source $\boldsymbol{n}_{\mathrm{RX2}}(t)$. The noise source no longer models the error in measurement devices alone, but also the stochastic variation due to the omitted variables. Again, define a state,

$$\boldsymbol{x}_{\mathrm{RX2}}(t) = (\boldsymbol{u}_{\mathrm{RX2}}(t-h, t), z_{\mathrm{RX2}}(t-h, t-1)), \tag{9}$$

where $h = \max[h_s - 1, h_c]$, and $\boldsymbol{x}_{\mathrm{RX2}}(t) \in X_{\mathrm{RX2}} = (\otimes_{i=1}^{h+1} U_{\mathrm{RX2}})(\otimes_{i=1}^{h} Z_{\mathrm{RX2}})$.

2.4 Level 1 receiver (RX1)

This is similar to the level 3 and 2 receivers with differences in the codebook mappings and notation. Denoising is,

$$z_{\mathrm{RX1}}(t) = \tilde{z}(t) - \boldsymbol{n}_{\mathrm{RX1}}(t), \tag{10}$$

where $\boldsymbol{n}_{\mathrm{RX1}}(t)\in Z$ and $z_{\mathrm{RX1}}(t)\in Z_{\mathrm{RX1}}$. The codebook implements deterministic mappings of form,

$$z_{\mathrm{RX1}}(t) : \boldsymbol{u}_{\mathrm{RX1}}(t) \mapsto z_{\mathrm{RX1}}(t). \tag{11}$$

Each codebook entry $z_{\mathrm{RX1}}(t)$ may be regarded as "typical" for its driver variables, in a similar manner to which the mean of a Gaussian distribution is typical of samples drawn from that Gaussian. The noise source $\boldsymbol{n}_{\mathrm{RX1}}(t)$ should now also describe the stochastic variation due to different initialisations and histories of driver variables. The present drivers form the state so $\boldsymbol{x}_{\mathrm{RX1}}(t)=\boldsymbol{u}_{\mathrm{RX1}}(t)$ where $\boldsymbol{x}_{\mathrm{RX1}}(t)\in X_{\mathrm{RX1}}=U_{\mathrm{RX1}}$.

2.5 Perfect image-model coupling

In this analysis, perfect image-model coupling is the transmission of a sequence of driver variables, without loss, via the true real-world process as encoder, the images as the transmitted message, and the proposed model as part of the decoder. Ideally $\boldsymbol{u}_{\mathrm{TX}}(t)=\boldsymbol{u}_{\mathrm{RX}q}(t)$. For the example of the level 3 receiver, assume that,

- the codebook mappings are identical, i.e. $z_{\mathrm{RX3}}^{-1}(t)\circ z_{\mathrm{TX}}(t)=I$, where I is the identity map,
- the domain of the receiver and transmitter codebooks are identically descriptive, i.e. $U_{\mathrm{RX3}}=U_{\mathrm{TX}}$,
- the transmitter noise source is correctly modelled by the receiver noise source, i.e. $P_{\mathrm{RX3}}(\boldsymbol{n}_{\mathrm{RX3}}(t))=P_{\mathrm{TX}}(\boldsymbol{n}_{\mathrm{TX}}(t)),\ \forall t$, and,
- processes and models implicit in the transmitter and receiver are assumed stationary.

The first two conditions assume the proposed codebook is correct, the third that the noise model is correct. Unfortunately, even under these conditions where the distribution of noise $\boldsymbol{n}_{\mathrm{RX3}}(t)$ is correct, the particular sample drawn from that noise distribution remains unknown. For this reason, even in the case of the level 3 receiver, perfect coupling may not be achievable. In many cases, the driver variables $\boldsymbol{u}_{\mathrm{TX}}(t)$ can only be recovered in the sense of maximum a-posteriori (MAP) or other estimates. For either of the level 3, 2 or 1 receivers, perfect coupling would only be possible if the transmitter and receiver were identical, and either noise samples were transmitted independently between the transmitter and receiver along a separate noiseless channel, or the supports of the noise distributions were always strictly less than the distances between neighbouring entries in the ranges of the transmitter and receiver codebooks. Decoding is described more fully in Sect. 3.

3 Receiver

The purpose of the receiver is to decode the image sequence $\tilde{z}(1,T)$ as a sequence of driver variables. Assume the receiver at level q implements this by first decoding the most appropriate state sequence through minimising a scalar objective function,

$$\hat{\boldsymbol{x}}_{\mathrm{RX}q}(1,T)(\tilde{z}(1,T))=\arg_{\boldsymbol{x}_{\mathrm{RX}q}(1,T)\in\otimes_{t=1}^{T}X_{\mathrm{RX}q}}\min f_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(1,T),\tilde{z}(1,T)), \tag{12}$$

subject to the constraint that there exists an underlying consistent driver sequence $\hat{\boldsymbol{u}}_{\mathrm{RX}q}(1,T)(\tilde{z}(1,T))$, i.e. an extraction mapping should exist,

$$\hat{\boldsymbol{x}}_{\mathrm{RX}q}(1,T)(\tilde{z}(1,T))\mapsto\hat{\boldsymbol{u}}_{\mathrm{RX}q}(1,T)(\tilde{z}(1,T)). \tag{13}$$

This section explains how the receiver achieves this. The definition of the receiver requires the specification of (1) a codebook, both the mapping as implemented by a deterministic model and its domain, (2) a noise model, and (3) a state transition model. The decoder attempts to find the best match between each image and a member of the codebook. This requires careful selection of (4) the objective function to measure the "goodness of match", and (5) a search mechanism, often heuristic, to navigate the codebook and find the member with maximum "goodness of match". These components are described in the remainder of this section.

3.1 Codebook

For the level q receiver, the codebook may be viewed as the set of deterministic mappings,

$$\mathcal{C}_{\mathrm{RX}q}=\{\boldsymbol{x}_{\mathrm{RX}q}(t)\mapsto z_{\mathrm{RX}q}(t),\ \forall\boldsymbol{x}_{\mathrm{RX}q}(t)\in X_{\mathrm{RX}q}\}. \tag{14}$$

These may be implemented using lookup tables, but more typically by empirical or mathematical models. The codebook domain $X_{\mathrm{RX}q}$ may also be constrained by the choice of model. For example an empirical model may impose lower and upper bounds on its driver variables, which in turn constrain the codebook domain. The domain may be quantised coarsely, or finely at machine precision. In ionospheric modelling, some codebooks may be more suitable for different tasks than others, e.g. some deterministic models may be better at describing high latitude processes while others may be more suitable for low latitude processes.

3.2 Noise model

Since the codebook is deterministic, stochastic variability must be introduced via a supplementary noise model. For the level q receiver, the noise process may be fully specified via a set of probability mass functions (PMFs),

$$\mathcal{N}_{\mathrm{RX}q}=\{P_{\mathrm{RX}q}(\boldsymbol{n}_{\mathrm{RX}q}(t)|\boldsymbol{x}_{\mathrm{RX}q}(t)),\forall\boldsymbol{n}_{\mathrm{RX}q}(t)\in Z,\forall\boldsymbol{x}_{\mathrm{RX}q}(t)\in X_{\mathrm{RX}q}\}. \tag{15}$$

In effect $\{\mathcal{C}_{\mathrm{RX}q},\mathcal{N}_{\mathrm{RX}q}\}$ defines a stochastic version of the deterministic codebook. Indeed if the model involved in image-model coupling is stochastic, then there is no need to define $\mathcal{C}_{\mathrm{RX}q}$ and $\mathcal{N}_{\mathrm{RX}q}$ separately. Ideally the codebook and noise model should replicate the stochastic variation in the transmitter (i.e. the real-world process). However this is a challenging task since considerable complexity is expected in the real-world variation, as described in Appendix A.

3.3 State transition model

Another supplementary model must be supplied governing the transitions between consecutive states. This may again be fully specified by a set of PMFs. For a level q receiver,

$$\mathcal{A}_{\mathrm{RX}q}=\{P_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(t)|\boldsymbol{x}_{\mathrm{RX}q}(t-1)),\forall\boldsymbol{x}_{\mathrm{RX}q}(t)\in X_{\mathrm{RX}q},\forall\boldsymbol{x}_{\mathrm{RX}q}(t-1)\in X_{\mathrm{RX}q}\}. \tag{16}$$

This state transition model must assign zero probability mass to those transitions which do not respect a temporally consistent sequence, e.g. for $\boldsymbol{u}_{\mathrm{RX}q}(1,T)$.

3.4 Objective function

The objective function should be derived from a decision theoretic perspective. For each sample $\tilde{z}(1,T)$, the decision rule should seek to minimise the conditional risk (Duda et al., 2001),

$$R(\boldsymbol{x}_{\mathrm{RX}q}(1,T)|\tilde{z}(1,T))=\sum_{\boldsymbol{y}_{\mathrm{RX}q}(1,T)\in\otimes_{i=1}^{T}X_{\mathrm{RX}q}} l(\boldsymbol{x}_{\mathrm{RX}q}(1,T),\boldsymbol{y}_{\mathrm{RX}q}(1,T))P_{\mathrm{RX}q}(\boldsymbol{y}_{\mathrm{RX}q}(1,T)|\tilde{z}(1,T)), \tag{17}$$

where $\boldsymbol{x}_{\mathrm{RX}q}(t)\in X_{\mathrm{RX}q},\forall t\in[1,T]$. The scalar function $l(\cdot,\cdot)$ is the loss, and the conditional risk is expressed as the average loss over a posterior distribution in the receiver. An intuitive choice of loss function is the squared L2 norm of the difference between the two arguments. However, under such a "regression-based" loss, the conditional risk is expensive to compute if samples must be drawn from the posterior. A simpler "classification-based" loss may instead be used (Duda et al., 2001),

$$l(\boldsymbol{x}_{\mathrm{RX}q}(1,T),\boldsymbol{y}_{\mathrm{RX}q}(1,T))=\begin{cases}0 \text{ if } \boldsymbol{x}_{\mathrm{RX}q}(1,T)=\boldsymbol{y}_{\mathrm{RX}q}(1,T)\\ 1 \quad \text{otherwise}\end{cases}. \tag{18}$$

This is preferable because it reduces computational cost since then,

$$R(\boldsymbol{x}_{\mathrm{RX}q}(1,T)|\tilde{z}(1,T))=1-P_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(1,T)|\tilde{z}(1,T)), \tag{19}$$

thereby avoiding the averaging operation. Hence an objective function may be expressed as follows, where logs are taken and unneccessary terms are discarded from the log posterior,

$$f_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(1,T),\tilde{z}(1,T))=-\ln P_{\mathrm{RX}q}(\tilde{z}(1,T)|\boldsymbol{x}_{\mathrm{RX}q}(1,T))-\ln P_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(1,T)). \tag{20}$$

The first of the two terms in the objective function models state-dependent noise, the second models memory in the state space (which includes memory in the driver source). Stochastic variation originates both in the channel and the source (e.g. see Khinchin, 1957). The first term is determined by the codebook $\mathcal{C}_{\mathrm{RX}q}$ and noise model $\mathcal{N}_{\mathrm{RX}q}$, the second by the state transition model $\mathcal{A}_{\mathrm{RX}q}$. The resultant decoder is the well-known maximum a-posteriori (MAP) decoder where the posterior acts as a measure of "goodness of match". The objective function may be compared with those derived in variational analysis in other applications such as meteorology (Daley, 1999). The remainder of this subsection considers how to obtain expressions for the objective function under different receivers, and details the assumptions implicit in decoding images using least sum square error optimisation.

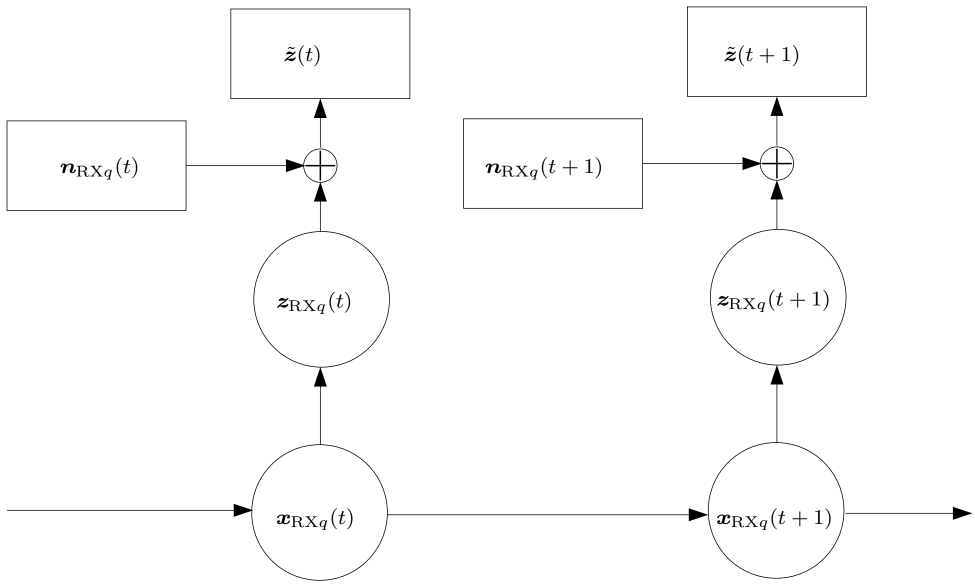

Fig. 2. Temporal portion of hidden Markov model (HMM) for modelling the receiver RXq, $q \in \{3, 2, 1\}$.

3.4.1 Discrete channel with memory

All receivers RXq, $q \in \{3, 2, 1\}$ may be modelled as hidden Markov processes of the form shown in Fig. 2, where the state is assumed to contain all information about the past which can influence the present and the future.

The noise process and state transition process are assumed stationary in time. The posterior is,

$$P_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(1,T)|\tilde{z}(1,T)) = \frac{\prod_{t=1}^{T} P_{\mathrm{RX}q}(\tilde{z}(t)|\boldsymbol{x}_{\mathrm{RX}q}(t))\, P_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(t)|\boldsymbol{x}_{\mathrm{RX}q}(t-1))}{P_{\mathrm{RX}q}(\tilde{z}(1,T))}, \tag{21}$$

where it is assumed that the state $\boldsymbol{x}_{\mathrm{RX}q}(0)$ is fully known, i.e. the history of driver variables and pre-noised images is known back to a timestep index of $-h$, if required. The objective function becomes,

$$f_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(1,T), \tilde{z}(1,T)) = -\sum_{t=1}^{T} \{\ln P_{\mathrm{RX}q}(\tilde{z}(t)|\boldsymbol{x}_{\mathrm{RX}q}(t)) + \ln P_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(t)|\boldsymbol{x}_{\mathrm{RX}q}(t-1))\}. \tag{22}$$

Both level 3 and level 2 receivers assume channel memory length and source memory length no greater than h and $h+1$ timesteps respectively. Of course, those memory lengths may in practice be reduced by constraining the PMFs in $\mathcal{N}_{\mathrm{RX}q}$ and $\mathcal{A}_{\mathrm{RX}q}$. For example, channel memory length may be reduced to zero in favour of source memory alone.

3.4.2 Discrete memoryless channel

The discrete memoryless channel is well known and is worthy of further consideration. Assume the probability of a state is independent of the previous state. This is not possible for a level 3 or 2 receiver. For the level 1 receiver,

$$f_{\mathrm{RX1}}(\boldsymbol{u}_{\mathrm{RX1}}(1,T), \tilde{z}(1,T)) = -\sum_{t=1}^{T} \{\ln P_{\mathrm{RX1}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{RX1}}(t)) + \ln P_{\mathrm{RX1}}(\boldsymbol{u}_{\mathrm{RX1}}(t))\}. \tag{23}$$

The noise is still state-dependent though it is no longer Markov. This channel assumes there is no source memory, i.e. there is independence between successive values of driver variables. This independence assumption is severe and most probably unrealistic for real-world systems. Driver variables naturally change smoothly, albeit at a certain level of scale. For the ionosphere, on all but the shortest time scales, any measurement of incident solar radiation may appear discontinous during the arrival of a solar flare. However for the majority of time, this measurement varies smoothly between consecutive timesteps, and the state transition model should favour such smooth changes. Also the assumption that the image is conditionally independent of previous variables given the current driver variables is unrealistic for systems such as the ionosphere where plasma accumulates with time. The ionosphere's response, as illustrated by its description, under the action of a given set of drivers, will vary depending on its past driver values. However the discrete memoryless assumption is convenient and, at the risk of reduced decoding accuracy, the likelihood $P_{\mathrm{RX1}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{RX1}}(t))$ may be regarded as averaged over all possible histories.

3.4.3 Discrete memoryless channel with state-independent Gaussian noise

The objective function for the discrete memoryless channel, as given in Eq. (23), may be simplified to yield the well-known sum square error and sum weighted square error objective functions. It is useful to consider the additional assumptions.

First, it is sometimes convenient to assume that the mapping $\boldsymbol{u}_{\mathrm{RX1}}(t) \mapsto z_{\mathrm{RX1}}(t)$ is injective $\forall t \in [1, T]$ (e.g. see "Mappings" by I. N. Sneddon in Sneddon, 1976). The injective assumption may not be unreasonable, particularly when a vector of few driver variables maps into a large image with many components. Then the objective function in Eq. (23) becomes,

$$f'_{\mathrm{RX1}}(\boldsymbol{u}_{\mathrm{RX1}}(1,T), \tilde{z}(1,T)) = -\sum_{t=1}^{T} \{\ln P_{\mathrm{RX1}}(\tilde{z}(t)|z_{\mathrm{RX1}}(t)) + \ln P_{\mathrm{RX1}}(\boldsymbol{u}_{\mathrm{RX1}}(t))\}. \quad (24)$$

The prior $P_{\mathrm{RX1}}(\boldsymbol{u}_{\mathrm{RX1}}(t))$ should reflect the frequency of occurrence of different values of driver variables over all possible histories. For example, for the ionosphere the prior may be calculated using measurements collected over a full solar cycle.

Next consider the noise process is state-independent and is stationary in time. Hence $P_{\mathrm{RX1}}(\tilde{z}(t)|z_{\mathrm{RX1}}(t)) = \delta(\tilde{z}(t) - z_{\mathrm{RX1}}(t), \boldsymbol{n}_{\mathrm{RX1}}) P_{\mathrm{RX1}}(\boldsymbol{n}_{\mathrm{RX1}})$, where $\delta(\cdot,\cdot)$ is the Kronecker delta, simplifying the noise model $\mathcal{N}_{\mathrm{RX1}}$ considerably to the specification of a single PMF. This is unlikely for systems such as the ionosphere. Additionally, assume the noise model is a zero-mean discretised Gaussian so $P_{\mathrm{RX1}}(\boldsymbol{n}_{\mathrm{RX1}}) = N(\boldsymbol{n}_{\mathrm{RX1}}; \boldsymbol{0}, \mathbf{R})$ where $\mathbf{R}$ is the covariance matrix[3], and hence $P_{\mathrm{RX1}}(\tilde{z}(t)|z_{\mathrm{RX1}}(t)) = N(\tilde{z}(t); z_{\mathrm{RX1}}(t), \mathbf{R})$. Unfortunately, Gaussianity is unlikely for those description variables, for example line integrals of electron content in the case of the ionosphere, which are naturally nonnegative.

Finally, assume there is no prior preference in the driver variables so there is a uniform prior $P_{\mathrm{RX1}}(\boldsymbol{u}_{\mathrm{RX1}}(t))$ over U_{RX1}. Again this is probably unreasonable for many real-world systems. For the case of the ionosphere, quiet space weather occurs much more frequently than stormy space weather, and the prior over driver variables should reflect this. The objective function for the discrete memoryless channel in Eq. (24) may now be simplified to the following, where irrelevant terms have been discarded,

$$f''_{\mathrm{RX1}}(\boldsymbol{u}_{\mathrm{RX1}}(1,T), \tilde{z}(1,T)) = \sum_{t=1}^{T} z_{\mathrm{RX1}}(t)^{\top} \mathbf{R}^{-1} (\frac{1}{2} z_{\mathrm{RX1}}(t) - \tilde{z}(t)). \quad (25)$$

The more negative the objective function, the better is the "goodness of match". If the covariance matrix $\mathbf{R}$ is assumed diagonal, minimisation is equivalent to the conventional least sum weighted square error solution. The diagonal elements, i.e. weights, represent the relative "importance" of each component in the image. If each component is of equal importance, $\mathbf{R}$ may be set to Identity and the minimisation yields the least sum square error solution.

3.5 Search

Minimisation of the objective function over the full domain $\otimes_{t=1}^{T} X_{\mathrm{RX}q}$ of the receiver codebook requires a search mechanism. The simplest approach is full grid search. However computational cost increases with the number of state variables, and coarser grid searches must often be introduced. Otherwise conventional derivative-based optimisation is recommended. However if the codebook mapping is implemented by some deterministic empirical model or when a mathematical model is very complicated, the derivative of an objective function of type $f_{\mathrm{RX}q}(\cdot)$ may not be known analytically. In this case, derivative-free optimisation techniques may then be required. Alternative approaches include numerical approximation of gradients (e.g. see the algorithms in Powell, 2007) and sampling-based methods. An example of a sampling scheme is simulated annealing (Salamon et al., 2002) which, although it converges to a global minimum, requires many evaluations of the objective function. A variant of simulated annealing is fast annealing (Salamon et al., 2002). Further comments on decoders are given in Sect. 6.

4 Evaluation

It is sometimes useful to compare different codebooks for image-model coupling. Codebooks should not be compared in isolation, but in the context of the accompanying noise model, state transition model, objective function and search mechanism.

[3]Estimation of the covariance matrix is not included in the variational problem, and must be performed apriori.

The error rate associated with each full receiver is the risk (Duda et al., 2001),

$$\mathcal{E}(\mathrm{RX}q) = \sum_{\tilde{z}(1,T)\in\otimes_{t=1}^{T} Z} R(\hat{\boldsymbol{x}}_{\mathrm{RX}q}(1,T)|\tilde{z}(1,T))P_{\mathrm{TX}}(\tilde{z}(1,T)), \tag{26}$$

with the conditional risk as given in Eq. (17) under the loss as given in Eq. (18). If each level q receiver is correct such that $\mathcal{M}_{\mathrm{RX}q}=\{\mathcal{C}_{\mathrm{RX}q}, \mathcal{N}_{\mathrm{RX}q}, \mathcal{A}_{\mathrm{RX}q}\}$ exactly replicates the statistical properties of the transmitter under the constraints imposed within each receiver, then $\mathcal{E}(\mathrm{RX3})\leq\mathcal{E}(\mathrm{RX2})\leq\mathcal{E}(\mathrm{RX1})$. This relationship does not necessarily hold for incorrect receivers. Indeed when all receivers are incorrect, simpler receivers can often prove more robust and give lower error rates in practice.

However the error rate penalises the decoding of variables, with values which differ from those in the transmitter, with the same loss independent of the degree of difference. This may be misleading. An alternative quantity is the mutual information between the sequence of true driver variables $\boldsymbol{u}_{\mathrm{TX}}(1,T)$ and the sequence of received driver variables $\boldsymbol{u}_{\mathrm{RX}q}(1,T)$. Perfect coupling, or lossless transmission, maximises this quantity. Any attempt to improve, learn or adapt incorrect models should aim to increase this quantity. Using MacKay (2004), the mutual information may be expressed as,

$$I(\boldsymbol{u}_{\mathrm{RX}q}(1,T);\boldsymbol{u}_{\mathrm{TX}}(1,T)) = H(\boldsymbol{u}_{\mathrm{TX}}(1,T)) - H(\boldsymbol{u}_{\mathrm{TX}}(1,T)|\boldsymbol{u}_{\mathrm{RX}q}(1,T)), \tag{27}$$

where the first and second terms on the right hand side are respectively source and conditional entropies. Since the source entropy is assumed fixed, the following simpler quantity may instead be used. Using MacKay (2004) for the expression for conditional entropy,

$$\begin{aligned} F(\mathcal{M}_{\mathrm{RX}q}) &= -H(\boldsymbol{u}_{\mathrm{TX}}(1,T)|\boldsymbol{u}_{\mathrm{RX}q}(1,T)) \\ &= \sum_{\boldsymbol{u}_{\mathrm{RX}q}(1,T)\in\otimes_{t=1}^{T} U_{\mathrm{RX}q}} P_{\mathrm{RX}q}(\boldsymbol{u}_{\mathrm{RX}q}(1,T)) \sum_{\boldsymbol{u}_{\mathrm{TX}}(1,T)\in\otimes_{t=1}^{T} U_{\mathrm{TX}}} P(\boldsymbol{u}_{\mathrm{TX}}(1,T)|\boldsymbol{u}_{\mathrm{RX}q}(1,T)) \ln P(\boldsymbol{u}_{\mathrm{TX}}(1,T)|\boldsymbol{u}_{\mathrm{RX}q}(1,T)), \end{aligned} \tag{28}$$

where,

$$-H(\boldsymbol{u}_{\mathrm{TX}}(1,T)) \leq F(\mathcal{M}_{\mathrm{RX}q}) \leq 0. \tag{29}$$

Note that some form of mutual information decoding would be more compatible with this comparison scheme, but MAP decoding simplifies the implementation of decoders. Rearranging and introducing the latent variables $\tilde{z}(1,T)$,

$$\begin{aligned} F(\mathcal{M}_{\mathrm{RX}q}) = &\sum_{\boldsymbol{u}_{\mathrm{RX}q}(1,T)\in\otimes_{t=1}^{T} U_{\mathrm{RX}q}} \sum_{\tilde{z}(1,T)\in\otimes_{t=1}^{T} Z} \sum_{\boldsymbol{u}_{\mathrm{TX}}(1,T)\in\otimes_{t=1}^{T} U_{\mathrm{TX}}} P(\boldsymbol{u}_{\mathrm{RX}q}(1,T),\tilde{z}(1,T),\boldsymbol{u}_{\mathrm{TX}}(1,T)) \\ &\ln P(\boldsymbol{u}_{\mathrm{TX}}(1,T)|\boldsymbol{u}_{\mathrm{RX}q}(1,T)). \end{aligned} \tag{30}$$

Making some conditional independence assumptions reasonable for our communication channel,

$$\begin{aligned} F(\mathcal{M}_{\mathrm{RX}q}) = &\sum_{\boldsymbol{u}_{\mathrm{RX}q}(1,T)\in\otimes_{t=1}^{T} U_{\mathrm{RX}q}} \sum_{\tilde{z}(1,T)\in\otimes_{t=1}^{T} Z} \sum_{\boldsymbol{u}_{\mathrm{TX}}(1,T)\in\otimes_{t=1}^{T} U_{\mathrm{TX}}} P_{\mathrm{RX}q}(\boldsymbol{u}_{\mathrm{RX}q}(1,T)|\tilde{z}(1,T)) \\ &P_{\mathrm{TX}}(\tilde{z}(1,T)|\boldsymbol{u}_{\mathrm{TX}}(1,T))P_{\mathrm{TX}}(\boldsymbol{u}_{\mathrm{TX}}(1,T)) \ln P(\boldsymbol{u}_{\mathrm{TX}}(1,T)|\boldsymbol{u}_{\mathrm{RX}q}(1,T)). \end{aligned} \tag{31}$$

Unfortunately it is difficult to estimate the distributions defined on the transmitter variables. However for the level 1 receiver, the quantity may be approximated by assuming the transmitter and receiver codebook mappings are injective (e.g. see "Mappings" by I. N. Sneddon in Sneddon, 1976) and the noise process $\boldsymbol{n}_{\mathrm{TX}}(t)$ is negligible. Then $\tilde{z}(t)\approx z_{\mathrm{TX}}(t), \forall t\in[1,T]$. Drawing ℓ samples of type $\boldsymbol{u}_{\mathrm{TX}}(1,T)$ according to the prior $P_{\mathrm{TX}}(\boldsymbol{u}_{\mathrm{TX}}(1,T))$ then,

$$F(\mathcal{M}_{\mathrm{RX1}}) \approx \frac{1}{\ell}\sum_{l=1}^{\ell} \sum_{z_{\mathrm{RX1}}(1,T)\in\otimes_{t=1}^{T} Z_{\mathrm{RX1}}} P_{\mathrm{RX1}}(z_{\mathrm{RX1}}(1,T)|\tilde{z}_l(1,T)) \ln P_{\mathrm{RX1}}(\tilde{z}_l(1,T)|z_{\mathrm{RX1}}(1,T)), \tag{32}$$

where $\tilde{z}_l(1,T)$ is injectively mapped through the transmitter codebook from the lth sample drawn according to $P_{\mathrm{TX}}(\boldsymbol{u}_{\mathrm{TX}}(1,T))$. Ignoring all posterior probability mass not located at the decoded solution,

$$F(\mathrm{RX1}) \approx \frac{1}{\ell}\sum_{l=1}^{\ell} P_{\mathrm{RX1}}(\hat{z}_{\mathrm{RX1}}(1,T)|\tilde{z}_l(1,T)) \ln P_{\mathrm{RX1}}(\tilde{z}_l(1,T)|\hat{z}_{\mathrm{RX1}}(1,T)), \tag{33}$$

which yields an expression for the whole receiver RX1. Since the expression is only dependent on the receiver, the quality of the estimate depends on the veracity of the receiver's components including its models and the effectiveness of its search algorithm. In the context of ionospheric modelling, each of the ℓ samples may be a sequence collected from a different day's data. When ℓ is small, alternative quantities which also penalise model complexity should be considered, for example stochastic complexity (e.g. see Grünwald et al., 2005). It is difficult to evaluate a receiver based on a single sample, i.e. when ℓ=1, though useful insight may still be possible.

Continuing with the level 1 receiver, as in Sect. 3.4 assume the noise process for the memoryless channel is state-independent, zero-mean discretised Gaussian and stationary in time. Then,

$$F(\mathrm{RX1}) \approx \frac{1}{\ell}\sum_{l=1}^{\ell}\Big[\prod_{t=1}^{T} P_{\mathrm{RX1}}(\hat{z}_{\mathrm{RX1}}(t)|\tilde{z}_l(t))\Big]\sum_{t=1}^{T}\ln P_{\mathrm{RX1}}(\tilde{z}_l(t)|\hat{z}_{\mathrm{RX1}}(t)), \tag{34}$$

where,

$$P_{\mathrm{RX1}}(\hat{z}_{\mathrm{RX1}}(t)|\tilde{z}_l(t)) = \frac{P_{\mathrm{RX1}}(\tilde{z}_l(t)|\hat{z}_{\mathrm{RX1}}(t))\,P_{\mathrm{RX1}}(\hat{z}_{\mathrm{RX1}}(t))}{\sum_{z_{\mathrm{RX1}}\in Z_{\mathrm{RX1}}} P_{\mathrm{RX1}}(\tilde{z}_l(t)|z_{\mathrm{RX1}})\,P_{\mathrm{RX1}}(z_{\mathrm{RX1}})}, \tag{35}$$

and $P_{\mathrm{RX1}}(\tilde{z}_l(t)|\hat{z}_{\mathrm{RX1}}(t))=N(\tilde{z}_l(t);\,\hat{z}_{\mathrm{RX1}}(t),\mathbf{R})$.

5 Sensitivity

Sensitivity of a sequence of true images $z_{\mathrm{TX}}(1,T)$ to the different state variables, including driver variables, in $\boldsymbol{x}_{\mathrm{TX}}(1,T)$ is of scientific interest. For example, during a geomagnetic storm, what drivers are most influential in producing the particular electron content patterns in the ionosphere? If the true real-world process is nonlinear, sensitivity must typically be evaluated at each sequence of state variables of interest. At the state sequence $\boldsymbol{y}_{\mathrm{TX}}(1,T)$, the Fisher information (see "Fisher information" in Wikipedia, access: October 2008) is,

$$J_{ij}(\boldsymbol{y}_{\mathrm{TX}}(1,T);\{\mathcal{C}_{\mathrm{TX}},\mathcal{A}_{\mathrm{TX}}\}) = \sum_{z_{\mathrm{TX}}(1,T)\in\otimes_{t=1}^{T}Z_{\mathrm{TX}}} \left.\frac{\delta^2 \ln P(z_{\mathrm{TX}}(1,T)|\boldsymbol{x}_{\mathrm{TX}}(1,T))}{\delta[\boldsymbol{x}_{\mathrm{TX}}(1,T)]^i\,\delta[\boldsymbol{x}_{\mathrm{TX}}(1,T)]^j}\right|_{\boldsymbol{x}_{\mathrm{TX}}(1,T)=\boldsymbol{y}_{\mathrm{TX}}(1,T)} P_{\mathrm{TX}}(z_{\mathrm{TX}}(1,T)|\boldsymbol{x}_{\mathrm{TX}}(1,T))|_{\boldsymbol{x}_{\mathrm{TX}}(1,T)=\boldsymbol{y}_{\mathrm{TX}}(1,T)}, \tag{36}$$

where $[\boldsymbol{x}_{\mathrm{TX}}(1,T)]^i$ is the ith component in the sequence of state variables, and $\{\mathcal{C}_{\mathrm{TX}},\mathcal{A}_{\mathrm{TX}}\}$ refers to the real-world process. Relating to drivers, δ denotes discrete differences but under the assumption that all values in the relevant component in $\boldsymbol{x}_{\mathrm{TX}}(1,T)$ are uniformly spaced. Since the true images are unknown,

$$\begin{aligned} J_{ij}(\boldsymbol{y}_{\mathrm{TX}}(1,T);\{\mathcal{C}_{\mathrm{TX}},\mathcal{A}_{\mathrm{TX}}\}) &\approx J_{ij}(\boldsymbol{y}_{\mathrm{TX}}(1,T);\mathcal{M}_{\mathrm{TX}}) \\ &= \sum_{\tilde{z}(1,T)\in\otimes_{t=1}^{T}Z} \left.\frac{\delta^2 \ln P(\tilde{z}(1,T)|\boldsymbol{x}_{\mathrm{TX}}(1,T))}{\delta[\boldsymbol{x}_{\mathrm{TX}}(1,T)]^i\,\delta[\boldsymbol{x}_{\mathrm{TX}}(1,T)]^j}\right|_{\boldsymbol{x}_{\mathrm{TX}}(1,T)=\boldsymbol{y}_{\mathrm{TX}}(1,T)} P_{\mathrm{TX}}(\tilde{z}(1,T)|\boldsymbol{x}_{\mathrm{TX}}(1,T))|_{\boldsymbol{x}_{\mathrm{TX}}(1,T)=\boldsymbol{y}_{\mathrm{TX}}(1,T)}. \end{aligned} \tag{37}$$

Unfortunately, the transmitter distributions are also unknown, but the Fisher information may be approximated at the receiver,

$$\begin{aligned} J_{ij}(\boldsymbol{y}_{\mathrm{TX}}(1,T);\mathcal{M}_{\mathrm{TX}}) &\approx J_{ij}(\boldsymbol{y}_{\mathrm{RX}q}(1,T);\mathcal{M}_{\mathrm{RX}q}) \\ &= \sum_{\tilde{z}(1,T)\in\otimes_{t=1}^{T}Z} \left.\frac{\delta^2 \ln P_{\mathrm{RX}q}(\tilde{z}(1,T)|\boldsymbol{x}_{\mathrm{RX}q}(1,T))}{\delta[\boldsymbol{x}_{\mathrm{RX}q}(1,T)]^i\,\delta[\boldsymbol{x}_{\mathrm{RX}q}(1,T)]^j}\right|_{\boldsymbol{x}_{\mathrm{RX}q}(1,T)=\boldsymbol{y}_{\mathrm{RX}q}(1,T)} P_{\mathrm{RX}q}(\tilde{z}(1,T)|\boldsymbol{x}_{\mathrm{RX}q}(1,T))|_{\boldsymbol{x}_{\mathrm{RX}q}(1,T)=\boldsymbol{y}_{\mathrm{RX}q}(1,T)}. \end{aligned} \tag{38}$$

The Fisher information defined upon the receiver is subject to the receiver's statistical assumptions, and may be a poor approximation to that defined on the transmitter. Since the Fisher information is defined on $\mathcal{M}_{\mathrm{RX}q}$, it does not depend on the optimisation algorithm. In terms of the objective function, and assuming $P_{\mathrm{RX}q}(\boldsymbol{x}_{\mathrm{RX}q}(1,T))$ is uniform,

$$J_{ij}(\boldsymbol{y}_{\mathrm{RX}q}(1,T);\mathcal{M}_{\mathrm{RX}q}) = -\sum_{\tilde{z}(1,T)\in\otimes_{t=1}^{T}Z} \left.\frac{\delta^2 f_{\mathrm{RX}q}(\tilde{z}(1,T),\boldsymbol{x}_{\mathrm{RX}q}(1,T))}{\delta[\boldsymbol{x}_{\mathrm{RX}q}(1,T)]^i\,\delta[\boldsymbol{x}_{\mathrm{RX}q}(1,T)]^j}\right|_{\boldsymbol{x}_{\mathrm{RX}q}(1,T)=\boldsymbol{y}_{\mathrm{RX}q}(1,T)} P_{\mathrm{RX}q}(\tilde{z}(1,T)|\boldsymbol{x}_{\mathrm{RX}q}(1,T))|_{\boldsymbol{x}_{\mathrm{RX}q}(1,T)=y_{\mathrm{RX}q}(1,T)}. \tag{39}$$

Given a single sample $\tilde{z}_l(1,T)$ and a level 1 receiver, a possible estimate at sequence $\boldsymbol{v}_{\mathrm{RX1}}(1,T)$ is,

$$J_{ij}(\boldsymbol{v}_{\mathrm{RX1}}(1,T);\mathcal{M}_{\mathrm{RX1}}) \approx -\frac{\delta^2 f_{\mathrm{RX1}}(\tilde{z}_l(1,T),\boldsymbol{u}_{\mathrm{RX1}}(1,T))}{\delta[\boldsymbol{u}_{\mathrm{RX1}}(1,T)]^i\,\delta[\boldsymbol{u}_{\mathrm{RX1}}(1,T)]^j}\Bigg|_{\boldsymbol{u}_{\mathrm{RX1}}(1,T)=\boldsymbol{v}_{\mathrm{RX1}}(1,T)}, \tag{40}$$

noting $\boldsymbol{x}_{\mathrm{RX1}}(1,T)=\boldsymbol{u}_{\mathrm{RX1}}(1,T)$. In this case the Fisher information is estimated using the negative Hessian of the objective function. However such an estimate may be misleading since it is "tuned" to one particular image sequence only, i.e. all the conditional probability mass for observed sequences is assumed located at $\tilde{z}_l(1,T)$.

6 Discussion

The selection of driver and image variables for $U_{\mathrm{RX}q}$ and Z respectively is critical. If the selection is not sufficiently descriptive, then the noise distributions implied in the transmitter may be very broad due to the effect of latent variables. Essentially, too much useful information may then be lost in the encoding process. Unfortunately, the choice of variables is often restricted by the particular receiver codebook, e.g. empirical or mathematical model, used.

In designing the receiver, the codebook is intuitively most important and effort should first be directed at improving this. The choice of noise model and state transition model should be data-dependent, since they should be learnt from data. This influences the choice of objective function and level of receiver. The simpler receivers should be more robust if there is a lack of good quality data. However it is important to understand the assumptions implicit in the simpler receivers, especially regarding the conditional independences. Of course, the level 3 and level 2 receivers reduce to the level 1 receiver if the level 1 assumptions hold in the transmitter. The choice of h_s, h_c and h may be driven by limitations in data rather than scientific knowledge. The inverse problem of decoding state variables $\hat{\boldsymbol{x}}_{\mathrm{RX}q}(1,T)$ for a noisy image sequence $\tilde{z}(1,T)$ has a unique solution providing the objective function has a single global minimum.

The receiver codebook has been defined using a single deterministic model. However multiple models may be used in parallel where the data fusion occurs at the level of $\boldsymbol{z}_{\mathrm{RX}q}(1,T)$ or $\boldsymbol{x}_{\mathrm{RX}q}(1,T)$. For ionospheric modelling, the fusion may additionally use geographic information where alternative ionospheric models are weighted differently at different global locations, for example according to their ability to model low latitude or polar/auroral processes.

The approach described above is limited. Our HMM enforces specific conditional dependencies; it is possible to apply different graphical models, e.g. Ihler et al. (2007). For example, the PMFs in our approach may be further constrained by incorporating spatial dependencies between variables, such as between pixels in an image, e.g. Burlina and Alajaji (1998); Link and Kallel (2000). When the channel is not memoryless, our approach also assumes additive Markov noise; other noise models are possible, for example semi-Markov noise (e.g. Chvosta and Reineker, 1999), or nonadditive noise. The approach is able to model intersymbol interference (ISI), as explained in Ferrari et al. (2005). It can also model the finite-state Markov channel (FSMC), e.g. Goldsmith and Varaiya (1996); Li and Collins (2007), where the driver variables form the system input. Implementing the decoder is challenging; a possible implementation is the Viterbi algorithm, e.g. Rabiner and Juang (1993); Kavčić and Moura (2000). However in practice, decoders may be too computationally expensive to implement unless the codebook mapping is computed apriori and stored in lookup tables, or the codebook mapping is very simple. Regarding decoders, it may be possible to incorporate concepts from JSCD such as interleavers, e.g. Li and Collins (2007), and the decoding algorithm, linked to Baum-Welch estimation, in Garcia-Frias and Villasenor (2001). Generally, it is difficult estimating PMFs with limited data, though concepts from speech recognition such as state and mixture tying and language model smoothing, e.g. Rabiner and Juang (1993), may prove beneficial. The accurate estimation of PMFs in the noise and state transition models is the main drawback of this approach. In practice, many simplifications may be required, in which case the receiver may degrade to a much simpler form. As detailed in Maraun et al. (2004), memory is often associated with the correlation function, and may be finite or infinite. It would be useful to analyse our approach by relating it to such a definition of memory.

Similar "receivers" have been developed in nonlinear and linear time series analysis. Examples from nonlinear time series analysis include the non-linear autoreggresive (NLAR) model (Chatfield, 2004), non-linear moving average (NLMA) model (Tong, 1990) and state-dependent model (SDM) (Priestley, 1988); however these examples assume there is no hidden state, equivalent to there being no channel noise. Also, Variable order Markov Models (VMMs) (Begleiter et al., 2004) model memory, but again there are no hidden states. The ARMA-filtered hidden Markov model (Michalek et al., 2000) may possibly be regarded as a special case of the level 2 receiver, at least in concept.

Besides the application to the ionosphere, the approach assuming memory may be useful for other geophysical systems which exhibit "sluggish responses", e.g. long-term climate modelling. As stated, a challenge for such systems is the selection of driver and image variables. For example, drivers for the ionosphere should include measures of solar and geomagnetic activity since it is known the sun and geomagnetic state of the Earth influence the distribution and movement of ionospheric plasma (Hargreaves, 2003). However if a 3-dimensional map of electron density measurements is used as the image, but the resolution is too coarse, the ability of the receiver to "recognise" small-scale structures is inhibited. Image-model coupling is simply an attempt to recognise or classify an image in terms of its driver variables.

7 Conclusions

Presented above is an information theoretic framework describing image-model coupling when the true-world system has memory. Examples of such systems include the ionosphere and other geophysical systems with a "sluggish response". A discrete channel model is used to help quantify the match between images and model output, and analyse the coupling. The approach is statistical in nature. It should be possible to harness any increased availability of data to derive objective functions which better reflect the spatial and temporal statistical relationships in the true underlying process, and thereby improve coupling as measured by decoding error rate. However for complex systems such as the ionosphere it is probably more beneficial to first direct effort at improving the accuracy of the proposed model (i.e. the codebook). It is hoped that the framework described above may encourage the further use of statistical and information-based "tools" in image-model coupling and its analysis. In general, image-model coupling may be used to help us better understand the underlying processes which produce the effects being imaged, but also better understand the limitations of our proposed models too.

Appendix A

Distributions implicit in the transmitter

The noisy image $\tilde{z}(t)$ may be regarded as sampled from a distribution, the functional form of which varies with the number of conditional variables. An expression for the fully marginalised distribution $P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t))$ may be derived under the following assumptions, consistent with the transmitter illustrated in Fig. 1.

- In $U_{\mathrm{TX}} \otimes U'_{\mathrm{TX}}$, each driver variable is linearly independent of all other driver variables.
- In $Z \otimes Z'$, each description variable is linearly independent of all other description variables.
- Measurement noise is stationary temporally and is state-independent. Hence $P_{\mathrm{TX}}(\boldsymbol{n}_{\mathrm{TX}}(t)|z_{\mathrm{TX}}(t)){=}P_{\mathrm{TX}}(\boldsymbol{n}_{\mathrm{TX}})\forall z_{\mathrm{TX}}(t), \forall t$.
- The pre-noised image $z_{\mathrm{TX}}(t)$ is fully determined by an initialisation h_c timesteps previous where $h_c{\in}\mathbb{N}_0$, and the history of driver variables since then, so,

$$(\boldsymbol{u}_{\mathrm{TX}}(t{-}h_c, t), \boldsymbol{u}'_{\mathrm{TX}}(t{-}h_c, t), z_{\mathrm{TX}}(t{-}h_c), z'_{\mathrm{TX}}(t{-}h_c)) \mapsto z_{\mathrm{TX}}(t). \tag{A1}$$

  If no such value of h_c exists, then a value of h_c is chosen such that the history of driver variables prior to timestep $(t{-}h_c)$ has no significant effect on the likelihood of the current description given the full description at $t{-}h_c$.

So,

$$P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c, t), \boldsymbol{u}'_{\mathrm{TX}}(t-h_c, t), z_{\mathrm{TX}}(t-h_c), z'_{\mathrm{TX}}(t-h_c)) = P_{\mathrm{TX}}(\boldsymbol{n}_{\mathrm{TX}})\delta(\boldsymbol{n}_{\mathrm{TX}}, \tilde{z}(t) - z_{\mathrm{TX}}(t)), \tag{A2}$$

where $\delta(\cdot,\cdot)$ is the Kronecker delta. Introducing redundant variables into the list of conditional variables,

$$\begin{aligned} &P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c, t), \boldsymbol{u}'_{\mathrm{TX}}(t-h_c, t), z_{\mathrm{TX}}(t-h_c, t), z'_{\mathrm{TX}}(t-h_c, t)) \\ &\qquad = P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c, t), \boldsymbol{u}'_{\mathrm{TX}}(t-h_c, t), z_{\mathrm{TX}}(t-h_c), z'_{\mathrm{TX}}(t-h_c)). \end{aligned} \tag{A3}$$

Marginalising over the unknown description and driver variables, and substituting from above,

$$
\begin{aligned}
&P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t), z_{\mathrm{TX}}(t-h_c,t))\\
&= \sum_{\boldsymbol{u}'_{\mathrm{TX}}(t-h_c,t)\in\otimes_{i=1}^{h_c+1}U'_{\mathrm{TX}}} \sum_{z'_{\mathrm{TX}}(t-h_c,t)\in\otimes_{i=1}^{h_c+1}Z'_{\mathrm{TX}}} P_{\mathrm{TX}}(\boldsymbol{u}'_{\mathrm{TX}}(t-h_c,t), z'_{\mathrm{TX}}(t-h_c,t))\\
&\quad P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t), \boldsymbol{u}'_{\mathrm{TX}}(t-h_c,t), z_{\mathrm{TX}}(t-h_c,t), z'_{\mathrm{TX}}(t-h_c,t))\\
&= \sum_{\boldsymbol{u}'_{\mathrm{TX}}(t-h_c,t)\in\otimes_{i=1}^{h_c+1}U'_{\mathrm{TX}}} \sum_{z'_{\mathrm{TX}}(t-h_c,t)\in\otimes_{i=1}^{h_c+1}Z'_{\mathrm{TX}}} P_{\mathrm{TX}}(\boldsymbol{n}_{\mathrm{TX}})\delta(\boldsymbol{n}_{\mathrm{TX}}, \tilde{z}(t)-z_{\mathrm{TX}}(t))\\
&\quad [\prod_{a=0}^{h_c-1} P_{\mathrm{TX}}(z'_{\mathrm{TX}}(t-a)|z'_{\mathrm{TX}}(t-h_c,t-a-1), \boldsymbol{u}'_{\mathrm{TX}}(t-h_c,t))]\\
&\quad P_{\mathrm{TX}}(z'_{\mathrm{TX}}(t-h_c)|\boldsymbol{u}'_{\mathrm{TX}}(t-h_c,t))[\prod_{a=0}^{h_c-1} P_{\mathrm{TX}}(\boldsymbol{u}'_{\mathrm{TX}}(t-a)|\boldsymbol{u}'_{\mathrm{TX}}(t-h_c,t-a-1))]P_{\mathrm{TX}}(\boldsymbol{u}'_{\mathrm{TX}}(t-h_c)).
\end{aligned}
\tag{A4}
$$

Then,

$$
\begin{aligned}
&P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t))\\
&= \sum_{\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t-1)\in\otimes_{i=1}^{h_c}U_{\mathrm{TX}}} \sum_{z_{\mathrm{TX}}(t-h_c,t)\in\otimes_{i=1}^{h_c+1}Z_{\mathrm{TX}}} P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t), z_{\mathrm{TX}}(t-h_c,t))\\
&\quad P_{\mathrm{TX}}(\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t-1), z_{\mathrm{TX}}(t-h_c,t))\\
&= \sum_{\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t-1)\in\otimes_{i=1}^{h_c}U_{\mathrm{TX}}} \sum_{z_{\mathrm{TX}}(t-h_c,t)\in\otimes_{i=1}^{h_c+1}Z_{\mathrm{TX}}} P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t), z_{\mathrm{TX}}(t-h_c,t))\\
&\quad [\prod_{a=1}^{h_c-1} P_{\mathrm{TX}}(z_{\mathrm{TX}}(t-a)|z_{\mathrm{TX}}(t-h_c,t-a-1), \boldsymbol{u}_{\mathrm{TX}}(t-h_c,t-1))]\\
&\quad P_{\mathrm{TX}}(z_{\mathrm{TX}}(t)|z_{\mathrm{TX}}(t-h_c,t-1), \boldsymbol{u}_{\mathrm{TX}}(t-h_c,t-1))P_{\mathrm{TX}}(z_{\mathrm{TX}}(t-h_c)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t-1))\\
&\quad [\prod_{a=1}^{h_c-1} P_{\mathrm{TX}}(\boldsymbol{u}_{\mathrm{TX}}(t-a)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t-a-1))]P_{\mathrm{TX}}(\boldsymbol{u}_{\mathrm{TX}}(t-h_c)),
\end{aligned}
\tag{A5}
$$

where $P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t-h_c,t), z_{\mathrm{TX}}(t-h_c,t))$ is as given in Eq. (A4) above. This expression gives an indication of the complexity implied in the distribution $P_{\mathrm{TX}}(\tilde{z}(t)|\boldsymbol{u}_{\mathrm{TX}}(t))$. Much of the complexity derives from the conditional probability terms introduced in marginalising over, or "averaging out", all possible histories. Of course the expression is simplified if the deeper dependencies do not exist, for example if h_c is small, or if the source memory length is much shorter than the channel memory length, i.e. $h_s \ll h_c$. If each receiver is correct in the sense described in Sect. 4, then the level 3, 2 and 1 receivers must respectively replicate the statistical distributions in Eqs. (A2), (A4) and (A5).

Acknowledgements. Approximately and in brief, N. D. Smith was primarily responsible for the development and writing; C. N. Mitchell gave the initial context and motivation for the work; all helped in proofreading and giving suggestions. N. D. Smith and C. J. Budd are grateful for being supported in BICS by EPSRC grant GR/S86525/01, and N. D. Smith is also grateful for financial support from the STFC. C. N. Mitchell thanks the EPSRC for financial support. N. D. Smith would like to thank individuals for useful comments or discussion, in particular Simon Harris for help with expanding states over time to ensure Markov properties, Dimitry Pokhotelov regarding ionospheric science, Merrilee Hurn, Gundula Behrens, Gary Bust and Massimo Materassi (ISC-CNR). The authors would also like to thank the referee for helpful suggestions and comments.

Edited by: J. Kurths
Reviewed by: one anonymous referee

References

Abramson, N.: Information Theory and Coding, McGraw-Hill Electronic Sciences Series, McGraw-Hill Book Company, Inc., New York, 1963.

Beaulieu, N.: The Evaluation of Error Probabilities for Intersymbol and Cochannel Interference, IEEE T. Commun., 39, 1740–1749, 1991.

Begleiter, R., El-Yaniv, R., and Yona, G.: On Prediction Using Variable Order Markov Models, J. Artif. Intell. Res., 22, 385–421, 2004.

Beyreuther, M. and Wassermann, J.: Continuous earthquake detection and classification using discrete Hidden Markov Models, Geophys. J. Int., 175, 1055–1066, 2008.

Burlina, P. and Alajaji, F.: An Error Resilient Scheme for Image Transmission over Noisy Channels with Memory, IEEE T. Image Process., 7, 593–600, 1998.

Chatfield, C.: The Analysis of Time Series: An Introduction, Texts in Statistical Science, Chapman & Hall/CRC, CRC Press LLC, Sixth edn., Boca Raton, Florida, 2004.

Chvosta, P. and Reineker, P.: Dynamics under the influence of semi-Markov noise, Physica A, 268, 103–120, 1999.

Daley, R.: Atmospheric Data Analysis, Cambridge atmospheric and space science series, Cambridge University Press, Cambridge, 1999.

Duane, G., Tribbia, J., and Weiss, J.: Synchronicity in predictive modelling: a new view of data assimilation, Nonlinear Proc. Geoph., 13, 601–612, 2006.

Duda, R., Hart, P., and Stork, D.: Pattern Classification, A Wiley-Interscience Publication, John Wiley & Sons,Inc., Second edn., New York, 2001.

Ferrari, G., Colavolpe, G., and Raheli, R.: A Unified Framework for Finite-Memory Detection, IEEE J. Sel. Area. Comm., 23, 1697–1706, 2005.

Garcia-Frias, J. and Villasenor, J.: Joint Turbo Decoding and Estimation of Hidden Markov Sources, IEEE J. Sel. Area. Comm., 19, 1671–1679, 2001.

Goldsmith, A. and Varaiya, P.: Capacity, Mutual Information, and Coding for Finite-State Markov Channels, IEEE T. Inform. Theory, 42, 868–886, 1996.

Grünwald, P., Myung, I., and Pitt, M., eds.: Advances in Minimum Description Length: Theory and Applications, Neural Information Processing Series, The MIT Press, Cambridge, Massachusetts, 2005.

Guo, J.-S., Shang, S.-P., Shi, J., Zhang, M., Luo, X., and Zheng, H.: Optimal assimilation for ionospheric weather – Theoretical aspect, Space Sci. Rev., 107, 229–250, 2003.

Hargreaves, J.: The solar-terrestrial environment, Cambridge atmospheric and space science series, Cambridge University Press, Cambridge, 2003.

Ihler, A., Kirshner, S., Ghil, M., Robertson, A., and Smyth, P.: Graphical models for statistical inference and data assimilation, Physica D, 230, 72–87, 2007.

Kavčić, A. and Moura, J.: The Viterbi Algorithm and Markov Noise Memory, IEEE T. Inform. Theory, 46, 291–301, 2000.

Khinchin, A.: Mathematical Foundations of Information Theory, Dover Publications, Inc., New York, (translated by R. A. Silverman and M. D. Friedman), 1957.

Korn, G. and Korn, T.: Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review, McGraw-Hill,Inc., second, enlarged and revised edn., New York, 1968.

Li, T. and Collins, O.: A Successive Decoding Strategy for Channels With Memory, IEEE T. Inform. Theory, 53, 628–646, 2007.

Link, R. and Kallel, S.: Markov Model Aided Decoding for Image Transmission Using Soft-Decision-Feedback, IEEE T. Image Process., 9, 190–196, 2000.

MacKay, D.: Information Theory, Inference and Learning Algorithms, Cambridge University Press, Cambridge, 2004.

Maraun, D., Rust, H. W., and Timmer, J.: Tempting long-memory – on the interpretation of DFA results, Nonlin. Processes Geophys., 11, 495–503, 2004, http://www.nonlin-processes-geophys.net/11/495/2004/.

Michalek, S., Wagner, M., and Timmer, J.: A New Approximate Likelihood Estimator for ARMA-Filtered Hidden Markov Models, IEEE T. Signal Proc., 48, 1537–1547, 2000.

Powell, M.: A view of algorithms for optimization without derivatives, Mathematics TODAY, 43, 170–174, 2007.

Priestley, M.: Non-linear and Non-stationary Time Series Analysis, Academic Press Limited, London, 1988.

Rabiner, L. and Juang, B.-H.: Fundamentals of Speech Recognition, Prentice Hall Signal Processing Series, PTR Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1993.

Salamon, P., Sibani, P., and Frost, R.: Facts, Conjectures, and Improvements for Simulated Annealing, SIAM Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2002.

Sneddon, I.: Encyclopaedic Dictionary of Mathematics for Engineers and Applied Scientists, Pergamon Press Ltd., 1976.

Tong, H.: Non-linear Time Series: A Dynamical System Approach, Oxford Statistical Science Series, Oxford University Press, Oxford, 1990.

Wikipedia: http://www.wikipedia.org, last access: October 2008.

Wikle, C. and Berliner, L.: A Bayesian tutorial for data assimilation, Physica D, 230, 1–16, 2007.

Wikle, C., Berliner, L., and Cressie, N.: Hierarchical Bayesian space-time models, Environ. Ecol. Stat., 5, 117–154, 1998.