University of Wollongong

Research Online

Centre for Statistical & Survey Methodology Working Paper Series

Faculty of Engineering and Information Sciences

2013

Environmental informatics: uncertainty quantification in the environmental sciences

Noel Cressie
University of Wollongong, ncressie@uow.edu.au

Follow this and additional works at: https://ro.uow.edu.au/cssmwp

Recommended Citation

Cressie, Noel, Environmental informatics: uncertainty quantification in the environmental sciences, Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 4-13, 2013, 21. https://ro.uow.edu.au/cssmwp/124

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

04-13

Environmental Informatics: Uncertainty Quantification in the Environmental Sciences

Noel Cressie

Copyright © 2013 by the National Institute for Applied Statistics Research Australia, UOW.
Work in progress, no part of this paper may be reproduced without permission from the Institute.

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Chapter 1

Environmental Informatics: Uncertainty Quantification in the Environmental Sciences

Noel Cressie[1]
Distinguished Professor
National Institute for Applied Statistics Research Australia (NIASRA)
School of Mathematics and Applied Statistics
University of Wollongong, NSW 2522
Australia
e-mail: ncressie@uow.edu.au

1.1 Introduction

This exposition of environmental informatics is an attempt to bring current thinking about uncertainty quantification to the environmental sciences. Environmental informatics is a term that I first heard being used by Bronwyn Harch of Australia's Commonwealth Scientific and Industrial Research Organisation to describe a research theme within her organisation. Just as bioinformatics has grown and includes biostatistics as a sub-discipline, environmental informatics, or EI, has the potential to be much broader than classical environmental statistics (e.g., [3]).

[1]I would like to thank Eddy Campbell for his comments on an earlier draft, Rui Wang for his help in preparing Figure 1.1, Emily Kang for her help in preparing Figure 1.2, and Andrew Holder for his help in preparing the manuscript. This research was partially supported by the NASA Program, NNH11ZDA001N-OCO2 (Science Team for the OCO-2 Mission).

Which came first, the hypothesis or the data? In EI, we start with environmental data, but we use them to reveal, quantify, and validate scientific hypotheses with a panoply of tools from statistics, mathematics, computing, and visualisation.

There is a realisation now in science, including the environmental sciences, that there is uncertainty in the data, the scientific models, and the parameters that govern these models. Quantifying that uncertainty can be approached in a number of ways. To some, it means smoothing the data to reveal interpretable patterns; to data miners, it often means looking for unusual data points in a sea of "big data"; and to statisticians, it means all of the above, using statistical modelling to address questions like, "Are the patterns real?" and "Unusual in relation to what?"

In the rest of this chapter, I shall develop a vision for EI around the belief that Statistics is the science of uncertainty, and that behind every good data-mining or machine-learning technique is an implied statistical model. Computing even something as simple as a sample mean and a sample variance can be linked back to the very simplest of statistical models with a location parameter and additive homoskedastic errors. The superb book by Hastie, Tibshirani, and Friedman [20] shows the fecundity of establishing and developing such links. EI is a young discipline, and I would like to see it develop in this modern and powerful way, with uncertainty quantification through Statistics at its core.

In what follows, I shall develop a framework that is fundamentally about environmental data and the processes that produced them. I shall be particularly concerned with big, incomplete, noisy datasets generated by processes that may be some combination of non-linear, multi-scale, non-Gaussian, multivariate, and spatio-temporal. I shall account for all the uncertainties coherently using hierarchical statistical modelling, or HM (e.g., [6]), which is based on a series of conditional-probability models. Finally, through loss functions that assign penalties as a function of how far away an estimate is from its estimand, I shall use a decision-theoretic framework (e.g., [5]) to give environmental policymakers a way to make rational decisions in the presence of uncertainty, based on competing risks (i.e., probabilities).

1.2 Hierarchical statistical modelling

The building blocks of HM are the data model, the (scientific) process model, and the parameter model. If Z represents the data, Y represents the process, and θ represents the parameters (e.g., measurement-error variance and reaction-diffusion coefficients), then the *data model* is:

$$[Z|Y,\theta], \tag{1.1}$$

the *process model* is:

$$[Y|\theta], \tag{1.2}$$

and the *parameter model* is:

$$[\theta], \tag{1.3}$$

where $[A|B,C]$ is generic notation for the conditional-probability distribution of the random quantity A given B and C.

A statistical approach represents the uncertainties coherently through the joint-probability distribution, $[Z,Y,\theta]$. Using the building blocks (1.1)–(1.3), we can write:

$$[Z,Y,\theta] = [Z|Y,\theta]\cdot[Y|\theta]\cdot[\theta]. \tag{1.4}$$

The definition of entropy of a random quantity A is $\mathrm{E}(\log[A])$; by re-writing (1.4) as:

$$\mathrm{E}(\log[Z,Y,\theta]) = \mathrm{E}(\log[Z|Y,\theta]) + \mathrm{E}(\log[Y|\theta]) + \mathrm{E}(\log[\theta]),$$

we can see that the joint entropy can be partitioned into data-model entropy, process-model entropy, and parameter-model entropy. This results in a "divide and conquer" strategy that emphasises where scientists can put effort into understanding the sources of uncertainty and into designing scientific studies that control (and perhaps minimise some of) the entropy components.

The process Y and the parameters θ are unknown, but the data Z are known. (Nevertheless, the observed Z is still thought of as one of many possible that could have been observed, with a distribution $[Z]$.) At the beginning of all statistical inference is a step that declares what to condition on, and I propose that EI follow the path of conditioning on what is known, namely Z. Then the conditional probability distribution of all the unknowns *given* Z is:

$$[Y,\theta|Z] = [Z,Y,\theta]/[Z] = [Z|Y,\theta]\cdot[Y|\theta]\cdot[\theta]/[Z], \tag{1.5}$$

where the first equality is known as Bayes' Theorem [4]; (1.5) is called the *posterior distribution*, and we call (1.1)–(1.3) a *Bayesian hierarchical model* (BHM). Notice that $[Z]$ on the right-hand side of (1.5) is a normalising term that ensures that the posterior distribution integrates (or sums) to 1.

There is an asymmetry associated with the role of Y and θ, since (1.2) very clearly emphasises that $[Y|\theta]$ is where the "science" resides. It is equally true that $[Y,\theta] = [\theta|Y]\cdot[Y]$, however probability models for $[\theta|Y]$ and $[Y]$ do not follow naturally from the way that uncertainties are manifested. The asymmetry emphasises Y as the first priority for inference. As a consequence, we define the *predictive distribution*, $[Y|Z]$, which can be obtained from (1.5) by marginalisation:

$$[Y|Z] = \int [Z|Y,\theta]\cdot[Y|\theta]\cdot[\theta]\,\mathrm{d}\theta/[Z]. \tag{1.6}$$

Then inference on Y is obtained from (1.6). While (1.5) and (1.6) are conceptually straightforward, in EI we may be trying to evaluate them in global spatial or spatio-temporal settings where Z might be on the order of Gb or Tb, and Y might be of a similar order. Thus, HM requires innovative conditional-probability modelling in (1.1)–(1.3), followed by innovative statistical computing in (1.5) and (1.6). Leading cases involve spatial data (e.g., [1, 9]) and spatio-temporal data (e.g., [13]). Examples of dynamical spatio-temporal HM are given in Chapter 9 of Cressie and Wikle [13], and we also connect the literature in data assimilation, ensemble forecasting, blind-source separation, and so forth to the HM paradigm.

1.3 Decision-making in the presence of uncertainty

Let $\widehat{Y}(Z)$ be one of many decisions about Y based on Z. Some decisions are better than others, which can be quantified through a (non-negative) loss function, $L(Y, \widehat{Y}(Z))$. The Bayes expected loss is $\mathrm{E}(L(Y, \widehat{Y}))$, and we minimise this with respect to $\widehat{Y}$. Then it is a consequence of decision theory (e.g., [5]) that the optimal decision is:

$$Y^*(Z) = \arg\inf_{\widehat{Y}} \Big\{ \mathrm{E}(L(Y, \widehat{Y})|Z) \Big\}, \tag{1.7}$$

where for some generic function $g(\cdot)$, the notation $\mathrm{E}(g(Y)|Z)$ is used to represent the conditional expectation of $g(Y)$ given Z.

Sometimes $\mathrm{E}(L(Y, \widehat{Y})|\theta)$ is called the risk, but I shall call it the expected loss; sometimes $\mathrm{E}(L(Y, \widehat{Y}))$ is called the Bayes risk, but see above where I have called it the Bayes expected loss. In what follows, I shall reserve the word *risk* to be synonymous with *probability*.

Now, if θ were known, only Y remains unknown, and HM involves just (1.1)–(1.2). Then Bayes' Theorem yields:

$$[Y|Z, \theta] = [Z|Y, \theta] \cdot [Y|\theta]/[Z|\theta]. \tag{1.8}$$

In this circumstance, (1.8) is both the posterior distribution and the predictive distribution; because of the special role of Y, I prefer to call it the predictive distribution. The analogue to (1.7) when θ is known is, straightforwardly,

$$Y^*(Z) = \arg\inf_{\widehat{Y}} \Big\{ \mathrm{E}\left(L(Y, \widehat{Y})|Z, \theta \right) \Big\}. \tag{1.9}$$

Clearly, $Y^*(Z)$ in (1.9) also depends on θ.

Using the terminology of Cressie and Wikle (2011), an empirical hierarchical model (EHM) results if an estimate $\hat{\theta}(Z)$, or $\hat{\theta}$ for short, is used in place of θ in (1.8): Inference on Y is then based on the *empirical predictive distribution*,

$$[Y|Z, \hat{\theta}] = [Z, Y, \hat{\theta}] \cdot [Y|\hat{\theta}]/[Z|\hat{\theta}], \tag{1.10}$$

which means that $\hat{\theta}$ is also used in place of θ in (1.9).

BHM inference from (1.5) and (1.6) is coherent in the sense that it emanates from the well defined joint-probability distribution (1.4). However, the BHM requires specification of the prior $[\theta]$, and it often consumes large computing resources to obtain (1.5) and (1.6). The EHM's inference from (1.10) can be much more computationally efficient, albeit with an empirical predictive distribution that has smaller variability than the BHM's predictive distribution (e.g., [39]). Bayes' Theorem applied to BHM or EHM for spatio-temporal data results in a typically very-high-dimensional predictive distribution, given by (1.6) or (1.10), respectively, whose computation requires dimension reduction (e.g.,

[2, 10, 11, 22, 25, 29, 31, 41, 42]) and statistical-computing algorithms such as EM (e.g., [30]), MCMC (e.g., [34]), and INLA [36].

In the last 20 years, methodological research in Statistics has seen a shift from mathematical statistics towards statistical computing. Deriving an analytical form for (1.6) or (1.10) is almost never possible, but being able to *sample* realisations from them often is. This shift in emphasis has enormous potential for EI.

For economy of exposition, I feature the BHM in the following discussion. First, if I can sample from the posterior distribution, $[Y, \theta|Z]$, I can automatically sample from the predictive distribution, $[Y|Z]$, by simply ignoring the θ's in the posterior sample of (Y, θ). This is called a *marginalisation property* of sampling. Now suppose there is scientific interest in a summary $g(Y)$ of Y (e.g., regional averages, or regional extremes). Then an *equivariance property* of sampling implies that samples from $[g(Y)|Z]$ are obtained by sampling from $[Y|Z]$ and simple evaluating each member of the sample at $g(\cdot)$. This equivariance property is enormously powerful, even more so when the sampling does not require knowledge of the normalising term $[Z]$ in (1.5). The best known statistical computing algorithm that samples from the posterior and predictive distributions is MCMC (e.g., [34]).

Which summary of the predictive distribution $[g(Y)|Z]$ will be used to estimate the scientifically interesting quantity $g(Y)$? Too often, the posterior mean,

$$\mathrm{E}(g(Y)|Z) = \int g(Y)[Y|Z]\,\mathrm{d}Y,$$

is chosen as a "convenient" estimator of $g(Y)$. This is an optimal estimator when the loss function is squared-error: $L(g(Y), \widehat{g}) = (\widehat{g} - g(Y))^2$; see, for example, Berger [5]. However, squared-error loss assumes equal consequences (i.e., loss) for under-estimation as for over-estimation. When a science or policy question is about extreme events, the squared-error loss function is strikingly inadequate, yet scientific inference based on the posterior mean is ubiquitous.

Even if squared-error loss were appropriate, it would be incorrect to compute $\mathrm{E}(Y|Z)$ and produce $g(\mathrm{E}(Y|Z))$ as an optimal estimate, unless $g(\cdot)$ is a linear functional of Y. However, this is also common in the scientific literature. Under squared-error loss, the optimal estimate is $\mathrm{E}(g(Y)|Z)$, which is defined above. Notice that aggregating over parts of Y defines a linear functional g, but that taking extrema over parts of Y results in a highly non-linear functional g. Consequently, the supremum/infimum of the optimal estimate of Y (i.e., $g(\mathrm{E}(Y|Z))$) is a severe under-estimate/over-estimate of the supremum/infimum of Y (i.e., $g(Y)$).

1.4 Smoothing the data

EI is fundamentally linked to environmental data and the questions that resulted in their collection. Questions are asked of the scientific process Y, and the data Z paint an imperfect and incomplete picture of Y. Often, the first tool that

comes to a scientist's hand is a "data smoother," which here I shall call f. Suppose one defines

$$\widetilde{Y} \equiv f(Z); \tag{1.11}$$

notice that f "de-noises" (i.e., filters out highly variable components) and "fills in" where there are missing data. The scientist might be tempted to think of $\widetilde{Y}$ as data coming directly from the process model, $[Y|\theta]$, and use classical statistical likelihoods based on $[Y = \widetilde{Y}|\theta]$ to fit θ and hence the model $[Y|\theta]$. But this paradigm is fundamentally incorrect; science should incorporate uncertainty using a different paradigm. Instead of (1.11), suppose I write

$$\widetilde{Z} \equiv f(Z). \tag{1.12}$$

While the difference between (1.11) and (1.12) seems simply notational, conceptually it is huge.

The smoothed *data* $\widetilde{Z}$ should be modelled according to $[\widetilde{Z}|Y, \theta]$, and the process Y can be incorporated into an HM through $[Y|\theta]$. Scientific inference then proceeds from $[Y|\widetilde{Z}]$ in a BHM according to (1.6) or from $[Y|\widetilde{Z}, \hat{\theta}]$ in an EHM according to (1.10). The definition given by (1.12) concentrates our attention on the role of data, processes, and parameters in an HM paradigm and, as a consequence, it puts uncertainty quantification on firm inferential foundations [13, Chapter 2].

Classical frequentist inference could also be implemented through a marginal model (i.e., the likelihood), $[\widetilde{Z}|\theta] = \int [\widetilde{Z}|Y, \theta] \cdot [Y|\theta] \, \mathrm{d}Y$, although this fact is often forgotten when likelihoods are formulated. As a consequence, these marginal models can be poorly formulated or unnecessarily complicated when they do not recognise the role of Y in the probability modelling.

1.5 EI for spatio-temporal data

This section of the chapter gives two examples from the environmental sciences to demonstrate the power of the statistical-modelling approach to uncertainty quantification in EI.

1.5.1 Satellite remote sensing

Satellite remote sensing instruments are remarkable in terms of their optical precision and their ability to deliver measurements under extreme conditions. Once the satellite has reached orbit, the instrument must function in a near vacuum with low-power requirements, sensing reflected light (in the case of a passive sensor) through a highly variable atmospheric column.

The specific example I shall discuss here is that of remote sensing of atmospheric CO_2, a greenhouse gas whose increase is having, and will have, a large effect on climate change. The global carbon cycle describes where carbon is stored and the movement of carbon between these reservoirs. The oceans and

vegetation/soil are examples of CO_2 *sinks*, and fires and anthropogenic emissions are examples of CO_2 *sources*; of the approximately 8 Gt per year that enters the atmosphere, about half is anthropogenic. About 4 Gt stays in the atmosphere and the other 4 Gt is absorbed roughly equally by the oceans and terrestrial processes. This global increase of approximately 4 Gt of atmospheric CO_2 per year is unsustainable in the long term.

It is of paramount importance to be able to characterise precisely where and when sinks (and sources) occur. Because of a lack of globally extensive, extremely precise, and very densely sampled CO_2 data, these are largely unknown. Once the spatial and temporal variability of the carbon cycle is understood, regional climate projections can be made, and rational mitigation/adaptation policies can be implemented.

Although the atmosphere mixes rapidly (compared to the oceans), there is a lot of spatial variability as a function of both surface location and (geopotential) height. There is also a lot of temporal variability at any given location, as is clear from the US National Oceanic and Atmospheric Administration's CO_2 daily measurements from their Mauna Loa (Hawaii) observatory. Hence, we define atmospheric CO_2 as a spatio-temporal process, $Y(\mathbf{s}; t)$, at spatial coordinates $\mathbf{s}$ and time t. Here, $\mathbf{s}$ consists of $(\text{lon}, \text{lat}) = (x, y)$ and geopotential height h, that belongs to the spatial domain D_s, the extent of the atmosphere around Earth; and t belongs to a temporal domain D_t (e.g., t might index days in a given month).

There are several remote sensing instruments that measure atmospheric CO_2 (e.g., NASA's AIRS instrument and the Japanese space agency's GOSAT instrument); to improve sensitivity to near-surface CO_2, NASA built the OCO-2 instrument. (The original OCO satellite failed to reach orbit in 2009.) It allows almost pinpoint spatial-locational accuracy (the instrument's footprint is 1.1×2.25 km), resulting in high global data densities during any given month. However, its small footprint results in quite a long repeat-cycle of 16 days, making it harder to capture daily temporal variability at high spatial resolution. I am a member of NASA's OCO-2 Science Team that is concerned with all components of the data-information-knowledge pyramid referred to below in Section 1.6.

The physics behind the CO_2 retrieval requires measurements of CO_2 in the so-called strong CO_2 and weak CO_2 bands of the spectrum, and of O_2 in the oxygen A-band [14]. The result is a data vector of radiances $\mathbf{Z}(x, y; t)$, where $(x, y) = (\text{lon}, \text{lat})$ is the spatial location on the geoid, $D_g \equiv (-180°, +180°) \times (-90°, +90°)$, of the atmospheric column from footprint to satellite; and t is the time interval (e.g., a day or a week) during which the measurements (i.e., radiances) for that column were taken, where t ranges over the period of interest D_t. This vector is several-thousand dimensional, and there are potentially many thousands of such vectors per time interval. Hence, datasets can be very large.

The data are "noisy" due to small imperfections in the instrument, ubiquitous detector noise, and the presence of aerosols and clouds in the column. After applying quality-control flags based on aerosol, cloud, albedo conditions, some data are declared unreliable and hence "missing." The ideal is to estimate the

(dry air mole fraction) CO_2 amount, $Y(x, y, h; t)$ in ppm, as h varies down the atmospheric column centred at (x, y), at time t. When the column is divided up into layers centred at geopotential heights $h_1, \ldots, h_K$, we may write:

$$\mathbf{Y}_0(x, y; t) \equiv (Y(x, y, h_1; t), \ldots, Y(x, y, h_K; t))', \tag{1.13}$$

as the scientific process (i.e., state) of interest. The dimension of the state vector (1.13) is 20 for OCO-2, although 40 or so additional state variables, $\mathbf{Y}_1(x, y; t)$, are incorporated into $\mathbf{Y}(x, y; t) \equiv (\mathbf{Y}_0(x, y; t)', \mathbf{Y}_1(x, y; t)')'$, from which the radiative-transfer relation can be modelled as:

$$\mathbf{Z}(x, y; t) = \mathbf{F}_\theta(\mathbf{Y}(x, y; t)) + \boldsymbol{\varepsilon}(x, y; t). \tag{1.14}$$

In (1.14), the functional form of $\mathbf{F}_\theta$ is known (approximately) from the physics, but typically it requires specification of parameters θ. If θ were known, (1.14) is simply the *data model*, $[\mathbf{Z}(x, y; t)|Y, \theta]$ on the right-hand side of (1.4). The *process model*, $[Y|\theta]$, on the right-hand side of (1.4) is the joint distribution of $Y \equiv \{\mathbf{Y}(x, y; t) : (x, y) \in D_g, t \in D_t\}$, whose individual multivariate distributions are specified by OCO-2 ATB Document [32] to be multivariate Gaussian with mean vectors and covariance matrices calculated from forecast fields produced by the European Centre for Medium-Range Weather Forecasting (ECMWF). However, this specification of the multivariate marginal distributions does not specify the joint distribution, $[Y|\theta]$. Furthermore, informed guesses are made for the parameters in θ. The predictive distribution is given by (1.8), but this is not computed; a summary is typically used (e.g., the predictive mode). For more details, see Crisp et al. [15] and O'Dell et al. [33]. Validation of the estimated CO_2 values is achieved through TCCON data from a globally sparse but carefully calibrated network of land-based, upward-looking CO_2 monitoring sites (e.g., [43]).

Ubiquitously in the literature on remote sensing retrievals (e.g., [35]), it is the *predictive mode* of $[\mathbf{Y}(x, y; t)|\mathbf{Z}(x, y; t), \theta]$ that is chosen as the optimal estimator of $\mathbf{Y}(x, y; t)$. The subsequent error analysis in that literature is then concerned with deriving the mean vector and the covariance matrix of this estimator assuming that $\mathbf{F}_\theta$ is a linear function of its state variables [8]. However, the atmosphere involves highly complex interactions, and the radiative transfer function is known to be highly non-linear.

In Cressie and Wang [12], we enhanced the linear approximation by including the quadratic term of a second-order Taylor-series approximation, and we calculated the non-linearity biases of retrievals of CO_2 that were obtained from data collected by the GOSAT satellite at the locations in Australia shown in Figure 1.1. For the six retrievals (i.e., predictive modes), we calculated the following biases of column-averaged CO_2, or XCO2 (in units of ppm): $0.86, 1.15, 0.19, 1.15, -0.78$, and 1.40. Biases of this order of magnitude are considered to be important, and hence a systematic error analysis of remote sensing retrievals should recognise the non-linearity in $\mathbf{F}_\theta$.

It is important to note here that the predictive distribution, $[\mathbf{Y}(x, y; t)|\mathbf{Z}(x, y; t), \theta]$, is different from the predictive distribution, $[\mathbf{Y}(x, y; t)|Z, \theta]$, and I propose that

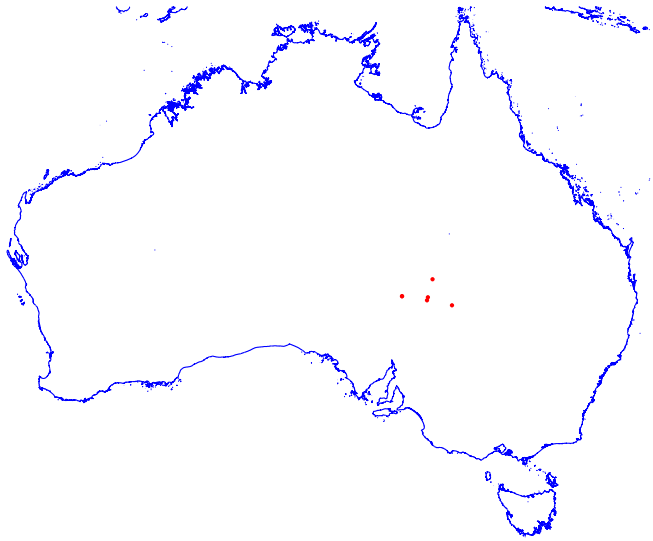

Figure 1.1: Locations of six GOSAT soundings where retrievals of XCO2 were obtained (between June 5, 2009 and July 26, 2009)

it is the latter that we should use when computing the optimal estimate of $\mathbf{Y}(x, y; t)$ from (1.9). This is based on the left-hand side of (1.8), which represents the "gold standard" to which all approximations should be compared. In practice, it would be difficult to obtain the predictive distribution, $[\mathbf{Y}(x, y; t)|Z, \theta]$, for every retrieval, so it makes sense to summarise it with its first two moments. In future research, I shall compare the linear approximations of Connor et al. [8] to the quadratic approximations of Cressie and Wang [12] by comparing them to the gold standard.

The mode should be considered to be just one possible summary of the predictive distribution; its corresponding loss function is

$$L(Y, \widehat{Y}) = \begin{cases} 0, & \text{if } Y = \widehat{Y} \\ 1, & \text{if } Y \neq \widehat{Y}; \end{cases}$$

see, for example, Berger [5]. I shall refer to this as the 0-1 loss function. That is, should even one element of the approximately 60-dimensional estimated state vector miss its target, a fixed loss is declared, no matter how close it is to the missed target. And the *same* fixed loss is declared when all or some the elements miss their targets, by a little or a lot. From this decision-theoretic point of view, the predictive mode looks to be an estimate that in this context is difficult to justify.

The next phase of the analysis considers the dry air mole fraction (in ppm) of CO_2 averaged through the column from Earth's surface to the satellite, which we saw is denoted as XCO2. Let $\mathbf{Y}_0^*(x, y; t)$ denote the predictive mode ob-

tained from (1.8), which is the optimal estimate given by (1.9) with the 0-1 loss function. Then XCO2$(x, y; t)$ is estimated by

$$\widehat{\text{X}}\text{CO2}(x, y; t) \equiv \mathbf{Y}_0^*(x, y; t)'\mathbf{w}, \tag{1.15}$$

where the weights $\mathbf{w}$ are given in OCO-2 ATB Document [32]. From this point of view, $\widehat{\text{X}}\text{CO2}(x, y; t)$ is the result of applying a smoother f to the raw radiances $\mathbf{Z}(x, y; t)$. The set of "retrieval data" over the time period D_t are $\{\widehat{\text{X}}\text{CO2}(x_i, y_i; t_i) : i = 1, \ldots, n\}$ given by (1.15), which we saw from (1.12) can be written as $\widetilde{Z}$; and Y is the multivariate spatio-temporal field $\{\mathbf{Y}(x, y; t) : (x, y) \in D_g, t \in D_t\}$, where recall that D_g is the geoid and the period of interest D_t might be a month, say.

The true column-averaged CO_2 field over the globe is a function of Y:

$$g_V(Y) \equiv \{\text{XCO2}(x, y; t) : (x, y) \in D_g, t \in D_t\}, \tag{1.16}$$

where the subscript V signifies *vertical averaging* of Y through the column of atmosphere from the satellite's footprint on the Earth's surface to the satellite. Then applying the principles set out in the previous sections, we need to construct spatio-temporal probability models for $[\widetilde{Z}|Y, \theta]$ and $[Y|\theta]$, and either a prior $[\theta]$ or an estimate $\hat{\theta}$ of θ. This will yield the predictive distribution of Y and hence that of $g_V(Y)$. Katzfuss and Cressie [25, 26] have implemented both the EHM where θ is estimated and the BHM where θ has a prior distribution, to obtain respectively, the empirical predictive distribution and the predictive distribution of $g_V(Y)$ based on $\widetilde{Z}$. The necessary computational efficiency is achieved by dimension reduction using the Spatio-Temporal Random Effects (STRE) model (e.g., [25]). Animated global maps of the predictive mean of $g_V(Y)$ using both approaches, based on AIRS CO_2 column averages, are shown in the SSES Web-Project, "Global Mapping of CO_2" (see Figure 2 at www.stat.osu.edu/∼sses/collab_co2.html). The regional and seasonal nature of CO_2 becomes obvious by looking at these maps. Uncertainty is quantified by the predictive standard deviations, and their heterogeneity (due to different atmospheric conditions and different sampling rates in different regions) is also apparent from the animated maps.

It is worth pointing out that the "smoothed" data, $\widetilde{Z} \equiv \{\widehat{\text{X}}\text{CO2}(x_i, y_i; t_i) : i = 1, \ldots, n\}$, are different from the original radiances, $Z \equiv \{\mathbf{Z}(x_i, y_i; t_i) : i = 1, \ldots, n\}$. Thus, $[Y|\widetilde{Z}, \theta]$ is different from $[Y|Z, \theta]$. Basing scientific inference on the latter, which contains all the data, is to be preferred, but practical considerations and tradition mean that the information-reduced, $\widetilde{Z} = f(Z)$, is used for problems such as flux estimation.

Since there is strong interest from the carbon-cycle-science community in regional *surface* fluxes, *horizontal* averaging should be a more interpretable summary of Y than vertical averaging. Let $g_1(\mathbf{Y}(x, y; t))$ denote the surface CO_2 concentration with units of mass/area. For example, this could be obtained by extrapolating the near-surface CO_2 information in $\mathbf{Y}_0(x, y; t)$. Then define

$$\overline{Y}(x, y; t) \equiv \textstyle\int_{R(x,y)} g_1(Y(u, v; t))\, \mathrm{d}u\mathrm{d}v \Big/ \int_{R(x,y)} \mathrm{d}u\mathrm{d}v,$$

and

$$g_H(Y) \equiv \{\overline{Y}(x,y;t) : (x,y) \in D_g, t \in D_t\}, \tag{1.17}$$

where the subscript H signifies horizontal averaging, and where $R(x,y)$ is a prespecified spatial process of areal regions on the geoid that defines the horizontal averaging. (It should be noted that R could also be made a function of t, and indeed it probably should change with season.) For a prespecified time increment τ, define

$$\Delta(x,y;t) \equiv \{\overline{Y}(x,y;t+\tau) - \overline{Y}(x,y;t)\}/\tau,$$

with units of mass/(area $\times$ time). Then the flux field is

$$g_F(Y) \equiv \{\Delta(x,y;t) : (x,y) \in D_g, t \in D_t\}. \tag{1.18}$$

At this juncture, it is critical that the *vector* of estimated CO_2 in the column, namely, $\mathbf{Y}_0^*(x_i,y_i;t_i)$, *replaces* $\widehat{\mathrm{X}}\mathrm{CO2}(x_i,y_i;t_i)$ to define the smoothed data, $\widetilde{Z}$. Then the data model $[\widetilde{Z}|Y,\theta]$ changes, but critically the spatio-temporal statistical model for $[Y|\theta]$ is the *same* as that used for vertical averaging. Recall the equivariance property that if Y is sampled from the predictive distribution (1.6) or (1.8), the corresponding samples from $g_H(Y)$ and $g_F(Y)$ yield their corresponding predictive distributions. The HM paradigm allows other data sources (e.g., *in situ* TCCON measurements, data from other remote sensing instruments) to be incorporated into $\widetilde{Z}$ seamlessly (e.g., [31]).

The choice of τ is at the granularity of D_t, and the choice of R depends on the question being asked and the roughness of Earth's surface relative to the question. In a classical bias-variance trade-off, one wants $R(x,y)$ to be large enough for $g_F(x,y;t)$ to capture the dominant variability and small enough that the flux in $R(x,y)$ is homogeneous.

Carbon-cycle science has accounted for much of the dynamics of CO_2, but the carbon budget has consistently shown there to be a missing sink (or sinks). The OCO-2 instrument, with its almost pinpoint accuracy and high sensitivity near Earth's surface, offers an unprecedented opportunity to accurately estimate the sinks. From that point of view, the parts of Y that are of interest are lower quantiles of $g_F(Y)$, along with the (lon, lat)-regions where those quantiles occur. In Section 1.6, I argue that these queries of the process $g_F(Y)$ can be formalised in terms of loss functions; Zhang et al. [45] give an illustration of this for decadal temperature changes over the Americas.

This different approach to flux estimation is centrally statistical, and it is based on a spatio-temporal model for $[Y|\theta]$. There is another approach, one that bases $[Y|\theta]$ on an atmospheric transport model to incorporate the physical movement of voxels in the atmosphere and, consequently, the physical movement of CO_2 (e.g., [7, 19, 21, 28]). Motivated by articles such as Gourdji et al. [19], I expect that the two approaches could be combined, creating a physical-statistical model.

When $[Y|\theta]$ is different, the predictive distribution given by (1.8) is different, and clearly when L in (1.9) is different, the optimal estimate given by (1.9) is

different. This opens up a whole new way of thinking about flux estimation and quantifying its uncertainty, which is something I am actively pursuing as part of the OCO-2 Science Team.

1.5.2 Regional climate change projections

Climate is not weather, the latter being something that interests us on daily basis. Generally speaking, climate is the empirical *distribution* of temperature, rainfall, air pressure, and other quantities over long time scales (30 years, say). The empirical mean (i.e., average) of the distribution is one possible summary, often used for monitoring trends, although empirical quantiles and extrema may often be more relevant summaries for natural-resource management. Regional climate models (RCMs) at fine scales of resolution (20–50 km) produce these empirical distributions over 30-year time periods and can allow decision-makers to project what environmental conditions will be like 50–60 years in the future.

Output from an RCM is obtained by discretising a series of differential equations, coding them efficiently, and running the programs on a fast computer. From that point of view, an RCM is deterministic, and there is nothing stochastic or uncertain about it. However, uncertainties in initial and boundary conditions, in forcing parameters, and in the approximate physics associated with the spatial and temporal discretisations (e.g., [17, 18, 44]), allow us to introduce a probability model for the output, from which we can address competing risks (i.e., probabilities) of different projected climate scenarios. The RCM output can certainly be summarised statistically; in particular, it can be mapped. There is a small literature on spatial statistical analyses of RCMs, particularly of output from the North American Regional Climate Change Assessment Program (NARCCAP), administered by NCAR in Boulder, Colorado [23, 24, 27, 38]. My work in this area has involved a collaboration with NCAR scientists.

Kang and Cressie [23] give a comprehensive statistical analysis of the 11,760 regions (50×50 km pixels) in North America, for projected *average* temperature change, projected out to the 30-year-averaging period, 2041–2070. The technical features of our article are: it is fully Bayesian; data dimension is reduced from the order of 100,000 down to the order of 100 through a Spatial Random Effects, or SRE, model [11]; seasonal variability is featured; and consensus climate-change projections are based on more than one RCM. Suppose that the quantity of scientific interest Y is temperature change in degrees Celsius by 2070, which is modelled statistically as,

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \mathbf{S}(\mathbf{s})'\boldsymbol{\eta} + \xi(\mathbf{s}); \quad \mathbf{s} \in \text{North America}, \tag{1.19}$$

where $\mu(\cdot)$ captures large-scale trends, and the other two terms on the right-hand side of (1.19) are Gaussian processes that represent, respectively, small-scale spatially dependent random effects and fine-scale spatially independent variability. The basis functions in $\mathbf{S}(\cdot)$ include 80 multi-resolutional bisquare functions and five indicator functions that capture physical features such as elevation and proximity to water bodies. This defines $[Y|\theta]$.

Importantly, the 30-year-average temperature change obtained from the NARCCAP output (i.e., the data, Z) is modelled as the sum of a spatial process Y and a spatial error term that in fact captures spatio-temporal interaction:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \varepsilon(\mathbf{s}); \quad \mathbf{s} \in \text{North America}, \tag{1.20}$$

where $\varepsilon(\cdot)$ is a Gaussian white-noise process with a variance parameter σ^2_ε. This defines $[Z|Y, \theta]$. The target for inference is the *spatial* climate-change process Y, which is "hidden" behind the *spatio-temporal* "noisy" process Z. A prior distribution, $[\theta]$, is put on θ, and (1.6) defines the predictive distribution.

Here, θ is made up of the vector of spatial-mean effects $\boldsymbol{\mu}$, $\text{cov}(\boldsymbol{\eta})$, $\text{var}(\xi(\cdot))$, and σ^2_ε, and the prior $[\theta]$ is specified in the Appendix of Kang and Cressie [23]. From (1.6), we can deduce the "red zone" of North America in Figure 1.2. There, with 97.5% probability calculated pixel-wise, any 50×50 km pixel's $Y(\mathbf{s})$ that is above a 2°C sustainability threshold is coloured red. Here, Y and θ together are over $100,000$ dimensional, but the computational algorithms based on dimension reduction in the SRE model do not "break."

Since MCMC samples are taken from a (more than) 100,000-dimensional posterior distribution, many such probability maps like Figure 1.2 can be produced. For example, there is great interest in extreme temperature changes, so let k denote a temperature-change threshold; define the spatial probability field,

$$\Pr\left(Y(\cdot) > k|Z\right); \quad k \geq 0. \tag{1.21}$$

As k increases in (1.21), the regions of North America that are particularly vulnerable to climate change stand out. Decision-makers can query the BHM where, for NARCCAP, the query might involve the projected temperature increase in the 50×50 km pixel containing Columbus, Ohio. Or, it might involve the projected temperature increase over the largely agricultural Olentangy River watershed (which contains Columbus). From (1.6), one can obtain the *probabilities* (i.e., risks) of various projected climate-change scenarios, which represent real knowledge when weighing up mitigation and adaptation strategies at the regional scale.

This HM approach to uncertainty quantification opens up many possibilities: Notice that the occurrence-or-not of the events referred to above can be written as $I(Y(\mathbf{s}_C) > k)$ and $I(\int_O Y(\mathbf{s})\,\mathrm{d}\mathbf{s} / \int_O \mathrm{d}\mathbf{s} > k)$, where $I(\cdot)$ is an indicator function, $\mathbf{s}_C$ is the pixel containing Columbus, and O is the set of all pixels in the Olentangy River watershed. Then *squared-error* loss implies that $\mathrm{E}(I(Y(\mathbf{s}_C) > k)|Z) = \Pr(Y(\mathbf{s}_C) > k|Z)$ given by (1.21) is optimal for estimating $I(Y(\mathbf{s}_C) > k)$. Critically, other loss functions would yield different optimal estimates, since (1.7) depends on the loss function L. A policy maker's question translated into a loss function yields a tailored answer to that question. Quite naturally in statistical decision theory, different questions are treated differently and result in different answers.

More critically, the average (climate change over 30 years) Y can be replaced with an extreme quantile, say the 0.95 quantile, which I denote here as $g^{(.95)}(Y)$; this hidden process corresponds to temperature change that could cause extreme

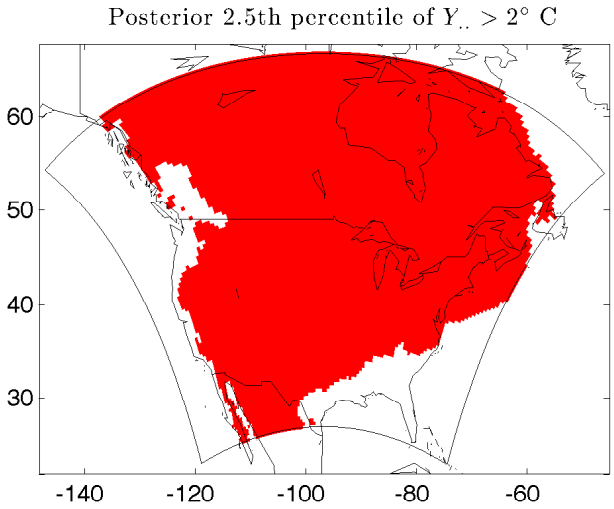

Figure 1.2: Regions of unsustainable (> 2°C with predictive probability 0.975) temperature increase obtained from pixel-wise predictive distributions, $[Y(\mathbf{s})|Z]$, where $\mathbf{s} \in$ North America.

stress to agricultural production in, for example, the Hunter Valley, NSW, Australia. Such projections for farmers in the "stressed" regions would be invaluable for planning crop varieties that are more conducive to higher temperature/lower rainfall conditions. That is, I propose making inference directly on extremal processes, and decisions should be made with loss functions that are tailor-made to the typical "what if" queries made by decision-makers.

Furthermore, output could have non-Gaussian distributions; for example, quantiles of temperature or rainfall would be skewed, for which spatial generalised linear models [16, 39] would be well suited: In this framework, (1.20) is replaced with the data model,

$$[Z(\mathbf{s}) = z|Y, \theta] = \text{Exp}(z; \text{E}(Z(\mathbf{s})|Y(\mathbf{s}), \theta)), \tag{1.22}$$

which are conditionally independent for pixels $\mathbf{s} \in$ North America. In (1.22), Exp denotes the one-parameter exponential family of probability distributions. Now consider a link function $\ell(\cdot)$ that satisfies, $\ell(\text{E}(Z(\mathbf{s})|Y(\mathbf{s}), \theta)) = Y(\mathbf{s})$; on this transformed scale, climate change $Y(\cdot)$ is modelled as the spatial process given by (1.19). In Sengupta and Cressie [39] and Sengupta et al. [40], we have developed spatial-statistical methodology for very large remote sensing datasets based on (1.22), that could be adapted to RCM projections. That methodology gives the predictive distribution, (1.10), which is summarised by mapping the predictive means, the predictive standard deviations (a measure of uncertainty), and the predictive extreme quantiles. Other data models could also be used in place of (1.20), such as the extreme-value distributions.

Increases in temperature generally lead to decreases in water availability, due to an increase in evaporation. By developing conditional-probability distributions of [Temperature] and [Rainfall|Temperature], we can infer the joint behaviour of [Rainfall, Temperature]. This is in contrast to the bivariate analysis in Sain et al. [37], and it is a further example of the utility of a conditional-probability modelling approach, here embedded in a multivariate hierarchical statistical model.

1.6 The knowledge pyramid

The knowledge pyramid has data at its base, information at its next tier, knowledge at its third tier, and decision-making at its apex. In the presence of uncertainty, I propose that EI have at its core the following steps: convert data into information by exploring the data for structure; convert information into knowledge by modelling the variability and inferring the etiology; and prepare the knowledge for decision-makers by translating queries into loss functions. These may not be the usual squared-error and 0-1 loss functions, which are often chosen for convenience. They may be asymmetric and multivariable, to reflect society's interest in extreme environmental events. Associated with each loss function (i.e., a query) is an optimal estimator (i.e., a wise answer) based on minimising the predictive expected loss; see (1.7) where the predictive risks (i.e., probabilities) and loss interact to yield an optimal estimator.

The societal consequences of environmental change, mitigation, and adaptation will lead to modelling of complex, multivariate processes in the social and environmental sciences. Difficult decisions by governments will involve choices between various mitigation and adaptation scenarios, and these choices can be made, based on the risks together with the losses that are built into EI's uncertainty quantification.

1.7 Conclusions

Environmental Informatics has an important role to play in quantifying uncertainty in the environmental sciences and giving policy-makers tools to make societal decisions. It uses data on the world around us to answer questions about how environmental processes interact and ultimately how they affect Earth's organisms (including *Homo sapiens*). As is the case for bioinformatics, environmental informatics not only requires tools from statistics and mathematics, but also from computing and visualisation. Although uncertainty in measurements and scientific theories mean that scientific conclusions are uncertain, a hierarchical statistical modelling approach gives a probability distribution on the set of all possibilities. Uncertainty is no reason for lack of action: Competing actions can be compared through competing Bayes expected losses.

The knowledge pyramid is a useful concept that data analysis, HM, optimal estimation, and decision theory can make concrete. Some science and policy

questions are very complex, so I am advocating an HM framework to capture the uncertainties and a series of queries (i.e., loss functions) about the scientific process to determine an appropriate course of action. Thus, a major challenge is to develop rich classes of loss functions that result in wise answers to important questions.

Bibliography

[1] S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, Boca Raton, FL, 2004.

[2] S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B*, 70:825–848, 2008.

[3] V. D. Barnett. *Environmental Statistics: Methods and Applications*. John Wiley and Sons, New York, NY, 2004.

[4] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

[5] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, NY, 1985.

[6] L. M. Berliner. Hierarchical Bayesian time-series models. In K. Hanson and R. Silver, editors, *Maximum Entropy and Bayesian Methods*, pages 15–22. Kluwer, Dordrecht, NL, 1996.

[7] F. Chevallier, F.-M. Bréon, and P. J. Rayner. Contribution of the Orbiting Carbon Observatory to the estimation of CO2 sources and sinks: Theoretical study in a variational data assimilation framework. *Journal of Geophysical Research*, 112(D09307), 2007. doi: 10.1029/2006JD007375.

[8] B. J. Connor, H. Boesch, G. Toon, B. Sen, C. Miller, and D. Crisp. Orbiting Carbon Observatory: Inverse method and prospective error analysis. *Journal of Geophysical Research: Atmospheres*, 113(D5), 2008. doi: 10.1029/2006JD008336.

[9] N. Cressie. *Statistics for Spatial Data*. John Wiley and Sons, New York, NY, rev. edition, 1993.

[10] N. Cressie and G. Johannesson. Spatial prediction for massive data sets. In *Australian Academy of Science Elizabeth and Frederick White Conference*, pages 1–11. Canberra, Australia: Australian Academy of Science, 2006.

[11] N. Cressie and G. Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B*, 70:209–226, 2008.

[12] N. Cressie and R. Wang. Statistical properties of the state obtained by solving a nonlinear multivariate inverse problem. *Applied Stochastic Models in Business and Industry*, 2013. doi: 10.1002/asmb.1946.

[13] N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. John Wiley and Sons, Hoboken, NJ, 2011.

[14] D. Crisp, D. J. Jacob, C. E. Miller, D. O'Brien, S. Pawson, J. T. Randerson, P. Rayner, R. J. Salawitch, S. P. Sander, B. Sen, G. L. Stephens, R. M. Atlas, P. P. Tans, G. C. Toon, P. O. Wennberg, S. C. Wofsy, Y. L. Yung, Z. Kuang, B. Chudasama, G. Sprague, B. Weiss, R. Pollock, F.-M. Bréon, D. Kenyon, S. Schroll, L. R. Brown, J. P. Burrows, P. Ciais, B. J. Connor, S. C. Doney, and I. Y. Fung. The Orbiting Carbon Observatory (OCO) mission. *Advances in Space Research*, 34:700–709, 2004.

[15] D. Crisp, N. M. Deutscher, A. Eldering, D. Griffith, M. Gunson, A. Kuze, L. Mandrake, J. McDuffie, J. Messerschmidt, C. E. Miller, I. Morino, B. M. Fisher, V. Natraj, J. Notholt, D. M. O'Brien, F. Oyafuso, I. Polonsky, J. Robinson, R. Salawitch, V. Sherlock, M. Smyth, H. Suto, C. O'Dell, T. E. Taylor, D. R. Thompson, P. O. Wennberg, D. Wunch, Y. L. Yung, C. Frankenberg, R. Basilio, H. Bosch, L. R. Brown, R. Castano, and B. Connor. The ACOS CO2 retrieval algorithm - Part II: Global X-CO2 data characterization. *Atmospheric Measurement Techniques*, 5:687–707, 2012.

[16] P. J. Diggle, J. A. Tawn, and R. A. Moyeed. Model-based geostatistics (with discussion). *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 47:299–350, 1998.

[17] J. P. Evans and S. Westra. Investigating the mechanisms of diurnal rainfall variability using a regional climate model. *Journal of Climate*, 25:7232–7247, 2012.

[18] M. J. Fennessy and J. Shukla. Seasonal prediction over North America with a regional model nested in a global model. *Journal of Climate*, 13: 2605–2627, 2000.

[19] S. M. Gourdji, K. L. Mueller, K. Schaefer, and A. M. Michalak. Global monthly averaged CO2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data. *Journal of Geophysical Research: Atmospheres*, 113(D21), 2008. doi: 10.1029/2007JD009733.

[20] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, 2nd edition, 2009.

[21] S. Houweling, F.-M. Bréon, I. Aben, C. Rödenbeck, M. Gloor, M. Heimann, and P. Ciasis. Inverse modeling of CO2 sources and sinks using satellite data: A synthetic inter-comparison of measurement techniques and their performance as a function of space and time. *Atmospheric Chemistry and Physics*, 4:523–538, 2004.

[22] E. L. Kang and N. Cressie. Bayesian inference for the Spatial Random Effects model. *Journal of the American Statistical Association*, 106:972–983, 2011.

[23] E. L. Kang and N. Cressie. Bayesian hierarchical ANOVA of regional climate-change projections from NARCCAP Phase II. *International Journal of Applied Earth Observation and Geoinformation*, 22:3–15, 2013.

[24] E.L. Kang, N. Cressie, and S. R. Sain. Combining outputs from the NARCCAP regional climate models using a Bayesian hierarchical model. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 61:291–313, 2012.

[25] M. Katzfuss and N. Cressie. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, 32: 430–446, 2011.

[26] M. Katzfuss and N. Cressie. Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics*, 23:94–107, 2012.

[27] C. G. Kaufman and S. R. Sain. Bayesian ANOVA modeling using Gaussian process prior distributions. *Bayesian Analysis*, 5:123–150, 2010.

[28] T. Lauvaux, A. E. Schuh, M. Uliasz, S. Richardson, N. Miles, A. E. Andrews, C. Sweeney, L. I. Diaz, D. Martins, P. B. Shepson, and K.J. Davis. Constraining the CO2 budget of the corn belt: Exploring uncertainties from the assumptions in a mesoscale inverse system. *Atmospheric Chemistry and Physics*, 12:337–354, 2012.

[29] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society, Series B*, 73:423–498, 2011.

[30] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, New York, NY, 2nd edition, 2008.

[31] H. Nguyen, N. Cressie, and A. Braverman. Spatial statistical data fusion for remote-sensing applications. *Journal of the American Statistical Association*, 107:1004–1018, 2012.

[32] OCO-2 ATB Document. OCO-2 level 2 full physics retrieval algorithm theoretical basis, 2010. URL `http://disc.sci.gsfc.nasa.gov/acdisc/documentation/OCO-2_L2_FP_ATBD_v1_rev4_Nov10.pdf`.

[33] C. W. O'Dell, B. Fisher, M. Gunson, J. McDuffie, C. E. Miller, V. Natraj, F. Oyafuso, I. Polonsky, M. Smyth, T. Taylor, GC Toon, B. Connor, P. O. Wennberg, D. Wunch, H. Bosch, D. O'Brien, C. Frankenberg, R. Castano, M. Christi, D. Crisp, and A. Eldering. The ACOS CO2 retrieval algorithm - Part I: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, 5:99–121, 2012.

[34] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, NY, 2nd edition, 2004.

[35] C. D. Rodgers. *Inverse Methods for Atmospheric Sounding*. World Scientific Publishing, Singapore, 2000.

[36] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71: 319–392, 2009.

[37] S. R. Sain, R. Furrer, and N. Cressie. Combining ensembles of regional climate model output via a multivariate Markov random field model. *Annals of Applied Statistics*, 5:150–175, 2011.

[38] E. S. Salazar, A. Finley, D. Hammerling, I. Steinsland, X. Wang, and P. Delamater. Comparing and blending regional climate model predictions for the American southwest. *Journal of Agricultural, Biological, and Environmental Statistics*, 16:586–605, 2011.

[39] A. Sengupta and N. Cressie. Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions. *Spatial Statistics*, 2013. doi: 10.1016/j.spasta.2013.02.002.

[40] A. Sengupta, N. Cressie, R. Frey, and B. Kahn. Statistical modeling of MODIS cloud data using the Spatial Random Effects model. In *2012 Proceedings of the Joint Statistical Meetings*, American Statistical Association, Alexandra, VA, pages 3111–3123, 2012.

[41] C. K. Wikle and N. Cressie. A dimension-reduced approach to space-time kalman filtering. *Biometrika*, 86:815–829, 1999.

[42] C. K. Wikle, R. F. Milliff, D. Nychka, and L. M. Berliner. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association*, 96:382–397, 2001.

[43] D. Wunch, P. Ahonen, S. C. Biraud, R. Castano, N. Cressie, D. Crisp, N. M. Deutscher, A. Eldering, M. L. Fisher, D. W. T. Griffith, M. Gunson, P. O. Wennberg, P. Heikkinen, G. Keppel-Aleks, E. Kyro, R. Lindenmaier, R. Macatangay, J. Mendonca, J. Messerschmidt, C. E. Miller, I. Morino, J. Notholt, G. C. Toon, F. A. Oyafuso, M. Rettinger, J. Robinson, C. M. Roehl, R. J. Salawitch, V. Sherlock, K. Strong, R. Sussmann, T. Tanaka,

D. R. Thompson, B. J. Connor, O. Uchino, T. Warneke, S. C. Wofsy, B. Fisher, G. B. Osterman, C. Frankenberg, L. Mandrake, and C. O'Dell. A method for evaluating bias in global measurements of CO2 total columns from space. *Atmospheric Chemistry and Physics*, 11:12317–12337, 2011.

[44] Y.K. Xue, R. Vasic, Z. Janjic, F. Mesinger, and K. E. Mitchell. Assessment of dynamic downscaling of the continental US regional climate using the Eta/SSiB regional climate model. *Journal of Climate*, 20:4172–4193, 2007.

[45] J. Zhang, P. F. Craigmile, and N. Cressie. Loss function approaches to predict a spatial quantile and its exceedance region. *Technometrics*, 50: 216–227, 2008.