Modelling Radionuclide Concentrations in Fucus on the Swedish West Coast

Applying Bayesian Distributed Lag Models to Radionuclide Transport

Pim Nelissen

Advised by Kristina Eriksson Stenström
Report for FYSP10: Advanced Research Project
7.5 ECTS, semester HT25

Department of Particle and Nuclear Physics
Lund University, Lund, Sweden
April 20, 2026

Abstract

Past and present measurements of radionuclide activity concentrations in the seaweed species Fucus *spp.* on the Swedish west coast have shown significant temporal variation. The current understanding is that marine discharges originating from remote nuclear facilities contribute to the observed activity levels. Initial statistical exploration has been done, but no detailed modelling has been performed thus far. In this project, an attempt is made to use a Bayesian dynamic generalized additive model (dGAM) to model observed activity concentrations at Särdal, Sweden as a response to remote discharges from facilities that are suspected to contribute to observed levels. The isotopes ^{14}C, ^{137}Cs, ^{129}I and ^{60}Co are studied due to their the highest data availability for this sampling location. For ^{129}I, variance is allowed between summer and winter samples in an effort to study seasonal effects on measured activity concentrations.

The models for ^{14}C do not follow literature findings of transport time or structure, hinting that the signal may be too weak, the model too complex, or key predictors (local NPP effluent) are missing from the model. The model for ^{137}Cs agrees reasonably well with literature, showing a dominant contribution from Baltic Sea outflow related to Chernobyl depositions, and initial effect from Sellafield, UK found after 4-5 years. The ^{129}I model also agrees reasonably with current understanding of transport time, with estimates for Sellafield to Särdal at around 4-5 years and La Hague to Särdal at 3-4 years. The ^{129}I model also hints at structurally higher activity concentration in the samples collected in winter, which corroborates recent findings. Finally, across the different isotope models we observe that the models tend to have very broad and dispersed effect well beyond the initial arrival of radioactive material, highlighting a non-trivial temporal pattern which the distributed lag approach is able to capture well.

We conclude that dGAMs are a powerful tool, but despite strong regularization, these models appear to be too flexible in the case where physics restrictions can reasonably be imposed. Future work could include more rigorous model validation, exploring explicit modelling of dilution factors, increasing the information in the model by sourcing higher frequency discharge data or exploring mixed-frequency models, and exploring distributed-lag approaches that allow physics restrictions to be imposed more directly than is possible in the current dGAM formulation.

Contents

1 Introduction

Significant temporal and spatial variation has been observed in radionuclide activity concentrations on the Swedish west coast, with contributions suspected from various sources. The seaweed species *Fucus serratus* (“toothed wrack") and *Fucus vesiculosus* (“bladderwrack"), collectively Fucus *spp.*, have been shown to be effective bioindicators for radionuclides in the marine environment of the Swedish west coast [1–5]. A collection of new and historical data on radionuclide concentrations from the Kattegat and the broader Swedish west coast have recently been presented by Mattsson et al. [1] and Stenström et al. [2]. The earliest samples date from 1967, which makes this a unique dataset for studying the long term effects of radionuclide discharges on the marine environment. Mattsson et al. [1] explored various anthropogenic radionuclides including Carbon-14, Caesium-137, Iodine-129 and Cobalt-60. Besides contributions from 1986 Chernobyl disaster for certain isotopes, the Barsebäck and Ringhals nuclear power plants have also been linked with elevated levels of radionuclides such as Carbon-14. Marine discharges from spent nuclear fuel reprocessing facilities such as those at Sellafield, United Kingdom and La Hague, France, have also been linked to the temporal and spatial variations observed in the Kattegat and Northern European waters in general [1–5].

The recent re-commitment of the Swedish government to nuclear energy [6] means there is a possibility of additional radionuclide discharge into the Swedish marine environment in the near future. To accurately map such future discharges, the current temporal and spatial variations need to be more deeply understood. Understanding and mapping these variations is also relevant from a public health perspective. This is because seaweed forms a key link in the marine food cycle, meaning radionuclides can end up in various marine life, including those consumed by humans.

Estimates of transport time between various sites and the Swedish west coast have been performed mainly by estimating the distance between peaks in discharge and corresponding peaks in activity concentration and by ocean transport studies. There are many environmental processes involved in the processes of radionuclide transport and uptake in Fucus *spp.*, which complicates the relationship between discharge and observed concentrations, making the relationship not necessarily linear. The objective of this project is to formulate and fit a model that relates radionuclide discharges from selected nuclear facilities to the observed activity concentrations in Fucus *spp.* samples obtained from Särdal, Sweden. The focus is on developing statistical machinery that can account for effects which are distributed over the temporal scale. The approach that has been identified as potentially suitable is the class of models known as dynamic generalized additive models.

2 Experimental background

This section describes in some detail the mechanisms of radionuclide uptake and release in Fucus. Then, the way various radionuclides enter the marine environment is discussed. The chapter closes with a description of the sampling procedure, from collection to measurement.

2.1 Sampling

Various research groups [1–5] have used Fucus *spp.* as a bioindicator for monitoring the trends in radionuclide concentrations in European waters. Radionuclides in the marine environment originate from a wide variety of sources, including nuclear power, industry and hospitals. Two particular reprocessing facilities for nuclear fuel, located at Sellafield, UK and Cap de La Hague, France, have been of particular interest due to their long history of significant liquid discharges. When it comes to the Nordics, the settlement of Särdal (56.76N, 12.63E), marked in Figure 1 by a red square, has been used as a sampling location for Fucus *spp.* since 1967 [1]. The rocky seabed in this location has ample growth of Fucus *spp.*, and its location on the coast of the Kattegat sea area means there is both inflow from the North Sea as well as outflow from the Baltic Sea, as highlighted in Figure 1. This makes it an suitable location for studying the transport of radionuclides from wider Northern European waters to

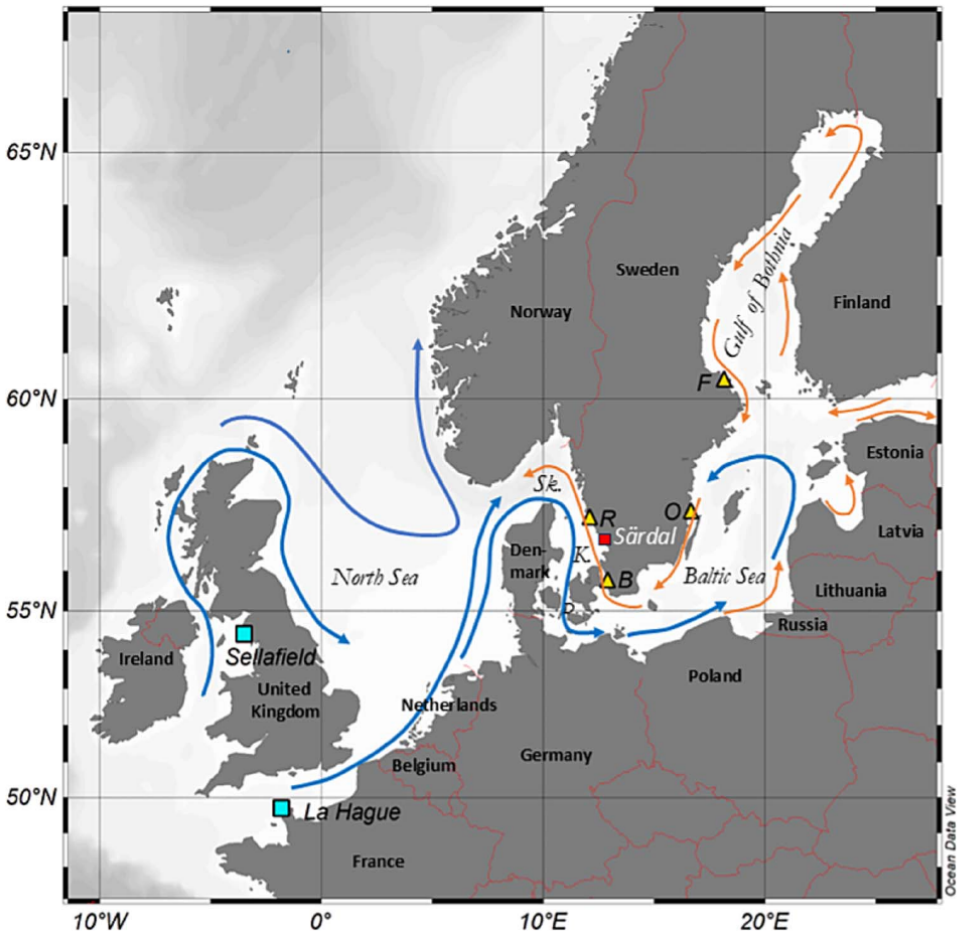


Figure 1: An overview of Northern European waters. The Särdal sampling site is highlighted in red, with major Swedish nuclear facilities denoted by yellow triangles. The North Sea and Baltic Sea currents are shown with blue and orange arrows respectively. Two nuclear fuel reprocessing facilities, at La Hague in France and Sellafield in the United Kingdom, are highlighted with blue squares. Figure reproduced from Mattsson et al. [1].

the Swedish west coast.

The samples have been collected by hand at depths of $0.3 - 0.9$m during various seasons. Care was taken in the selection of samples: several plants were obtained during every sampling occasion, with specimens that had other organisms growing on them were avoided [5]. The samples were dried overnight at $70-80^{\circ}$C, and ground into a powder and mixed. Measurement procedures differ between different isotopes – these processes are described in some more detail by Stenström and Mattsson [5].

2.2 Radionuclides in the environment of the Swedish west coast

Four isotopes (Carbon-14, Caesium-137, Iodine-129 and Cobalt-60) have been selected for this study, the choice of which will be motivated in Section 4.3. Here, the origins of these four radionuclides will be described, as well as environmental characteristics that are worth noting.

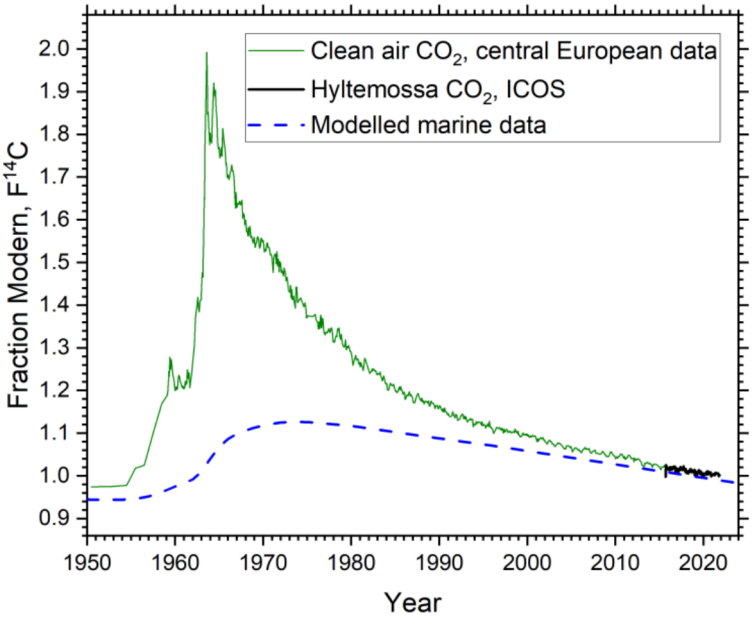


Figure 2: Atmospheric and modelled marine radiocarbon activity for the Northern Hemisphere, expressed in $F^{14}C$. The ^{14}C from atmospheric nuclear weapons tests of the 50s and 60s can be clearly seen as a visible atmospheric peak around 1963. The modelled marine data is for depths of 75 m. Figure reproduced from Stenström et al. [2].

Carbon-14 ($T_{1/2} = 5730$ years)

The radioactive isotope ^{14}C is naturally produced in the upper atmosphere by cosmic ray interaction with nitrogen in the n-p reaction

$$^{14}_{7}\mathrm{N} + \mathrm{n} \rightarrow ^{14}_{6}\mathrm{C} + \mathrm{p} + 0.626\ \mathrm{MeV}\,,$$

with an estimated 1.4×10^6 TBq/year entering the environment through this process [7]. The excess radiocarbon in the environment primarily originates from historic atmospheric nuclear weapons tests in the 1950s and 1960s, as can be seen in Figure 2. In 2008, UNSCEAR [8] estimated that a total of 350×10^6 TBq of ^{14}C has been added into the atmosphere from such tests. Civil applications of nuclear technology, such as nuclear power plants (NPPs), nuclear reprocessing facilities and certain other industries, also contribute to past and present radiocarbon concentrations in the environment. The historic public dose was dominated by Carbon-14 from the atmospheric nuclear weapons tests. In the present day, airborne releases of Carbon-14 from operational NPPs dominate the effective public dose in Sweden [2]. As for liquid discharges, Sellafield and La Hague have been identified as likely contributors to observed concentrations in Fucus *spp.* on the Swedish west coast. Marine discharge of Carbon-14 from Swedish NPPs are not included in effluent monitoring programmes, and thus there is thus no data available from these sites [2]. However, the releases from the Barsebäck NPP likely have an influence on ^{14}C levels around Särdal. Although the major water currents go North from Ringhals NPP (see Figure 1), releases from this plant could also have some influence on Särdal levels.

Atmospheric radiocarbon quickly oxidizes to CO_2, after which it enters the global carbon cycle. It can enter the marine environment in the form of dissolved inorganic carbon (DIC) through various pathways including photosynthesis in Fucus *spp.*. Excess ^{14}C can be quantified by the Fraction Modern, $F^{14}C$, where 1 $F^{14}C$ corresponds to the natural specific activity of carbon, around 226 Bq/kg C. Anthropogenic activity contributions would thus result in $F^{14}C > 1$.

The growth of Fucus *spp.* on the rocky seabed near the shore means atmospheric CO_2 enters the seaweed through photosynthesis [2]. Uptake also occurs through calcification processes in calcify-

ing species (e.g.*mollusca*), which use DIC as a carbon source. There are also various other so-called carbon pools through which DIC can enter and re-enter the marine environment around Fucus *spp.*, through processes such as the marine food cycle and sediment mixing [2, 9]. On top there are natural advection-diffusion processes in ocean environments that diffuse the DIC over time. All of this means that ^{14}C activity in Fucus *spp.* is not just affected by the uptake of 'fresh' atmospheric radiocarbon entering the marine environment for the first time. Rather, there is a continual recycling process along with non-trivial diffusion processes at play. All of this signals that the observed activity in Fucus *spp.* should be thought of as the result of various effects that occur at various temporal scales, rather than an instantaneous uptake.

Caesium-137 ($T_{1/2} = 30.08$ years)

The majority of ^{137}Cs in the Nordics today comes atmospheric releases from the 1986 Chernobyl disaster, with significant contributions from nuclear weapons testing as well. The deposition of this excess ^{137}Cs from Chernobyl has been a dominant contributor to observed activity history in the Baltic sea [10]. This is relevant to observations on the Swedish west coast, since the Kattegat receives significant volumes Baltic sea outflow [10–12]. Therefore, this source must be included in the analysis. However, contributions from the Sellafield discharges into the Irish sea are also significant enough to be considered in further study. Previous work has shown that sediment processes in the Irish sea can cause remobilisation of deposited radioactive caesium particulates [13]; this remobilisation is significant enough to be detected and should be considered in analysis of ^{137}Cs as a potential long-term but highly diffused source of activity on the Swedish west coast. Yearly discharges from La Hague are two orders of magnitude lower than Sellafield, and are unlikely to be a major contributor to observed levels in Fucus *spp.* sampled on the Swedish west coast.

Iodine-129 ($T_{1/2} = 1.57 \times 10^7$ years)

Radioactive iodine primarily originates from anthropogenic activities. It has been produced in excess as a fission by-product from both military and industrial activities, with the 1986 Chernobyl disaster also being a major source of ^{129}I. Unlike ^{137}Cs however, radioactive iodine originating from the Chernobyl releases is not believed to be a major contribution to the observed marine levels in the Nordics. Hou et al. [12] estimate that marine discharges from European facilities such as Sellafield and La Hague dominate the observed ^{129}I levels in the Kattegat, with Chernobyl release in Baltic sea outflow only accounting for about 0.1% of total levels in the Kattegat.

One well-studied but noteworthy characteristic of ^{129}I in Fucus *spp.* is the significant bioaccumulation that occurs. Furthermore, a significant seasonal difference has been observed, with ^{129}I activity concentrations in winter being significantly elevated compared to summer in the Fucus *spp.* samples collected at Särdal [1]. A recent study by Satoh, Wada, and Hisamatsu [14] on iodine concentrations in seagrass in Northern Japan seawaters showed a similar effect, and Nitschke et al. [15] found elevated levels of iodine in winter for both *Fucus vesiculosus* and *Fucus serratus*. The cause of this is not entirely clear as of today, but this is likely a combination of various environmental effects including metabolic processes and water temperature [15]. This seasonal effect will be explicitly included in the model.

Cobalt-60 ($T_{1/2} = 5.27$ years)

Like the other radionuclides discussed so far, the levels of ^{60}Co observed in Fucus *spp.* primarily originate from marine discharges of nuclear facilities. Mattsson et al. [1] describe that La Hague is the most likely explanation for the historically observed levels of ^{60}Co. Although shut down in 1990, the Winfrith NPP (UK, English Channel) is also noted as a potential source of ^{60}Co in the Kattegat area [1, 16]. Likewise, the Barsebäck and Ringhals NPPs, due to their proximity, likely have major influence.

3 Mathematical background

Most generally speaking, the goals of time series analysis are to forecast, to identify patterns (trends or seasonal structure) and to describe temporal relationships between quantities. What makes the anal-

ysis of time series challenging is the correlation between adjacent points; the ordering of data itself is relevant for modelling the underlying processes. This gives rise to a number of challenges. For example, many statistical methods assume that observations are *independent and identically distributed*, meaning the sampling is random [17]. This is not true for a time series, as samples are, due to their ordered nature, not independent, and often not normally distributed either [17]. The following sections will build up a time series model starting from a basic linear regression model. In Section 4, the machinery explained here will be applied to the radionuclide transport problem.

3.1 Linear regression for time series

Let y_t, $t = 1,\dots,N$ be the *dependent* time series of observations, which are predicted by *independent* time series $x_{t1}, x_{t2}, \dots, x_{tp}$. A basic regression model, which linearly links the predictors with the observed series, can then be formulated as

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_p x_{tp} + \varepsilon_i \,, \tag{1}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma)$ is an independent and identically distributed noise process. Equivalently, Equation 1 can be written as

$$Y_t | x_t \sim \mathcal{N}(\mu(x_t|\beta_0, \beta_1, \dots, \beta_p), \sigma);, \tag{2}$$

where $\mu(x_t|\beta_0, \beta_1, \dots, \beta_p) = E(Y|x_t, \beta_0, \beta_1, \dots, \beta_p)$. This means there is a linear relationship between the predictors and the expected value of the response. However, many real-life problems do not have a linear relationship between predictors and response [18]. To allow for a non-linear relation, one can introduce a smooth and monotonic *link function* $g(\mu)$ in the following way:

$$E[Y_t] = g^{-1}\left(\beta_0 + \beta_1 x_{t1} + \cdots + \beta_p x_{tp}\right) . \tag{3}$$

Equation 3 is called a *generalized linear model* (GLM). One can see now that the linear regression described earlier is simply a special case of a GLM with an identity link function.

3.2 Conceptually understanding distributed lag models

Equation 3 models the instantaneous effect of predictors x_{tp} on $E[Y_t]$. However, there are many processes imaginable were the effect of covariates is not instantaneous. In environmental or ecological processes, it is rather common that predictors have a delayed effect. Furthermore, the effect may be *distributed* over various time lags. This is not limited to only environmental processes; In 1965, Almon [19] was one of the first to describe a distributed lag model, which was in the context of econometrics. To give a practical example, a distributed lag model, describing the expected value of the dependent time series as a linear combination of historical values of covariates, could be

$$E[Y_t] = g^{-1}\left(\beta_0 + \sum_{l=0}^{L} \beta_l x_{t-l}\right) \tag{4}$$

where each parameter β_l quantifies how much x at time $t-l$ affects the current observed value $E[Y_t]$. The maximum lag L can be determined by knowledge of physical or otherwise process-related restraints, but will always be limited by the amount of data available. In effect only $N - L$ observations can be used for inference.

3.3 Generalized additive models using univariate smooths

Further generalization to the GLM can be achieved by replacing regression term $\beta_p x_{tp}$ with a smooth spline basis $f_p(x_{tp})$, which allows for more complex non-linear relationships to be modelled [18, 20]. For a single covariate, this may look like

$$E[Y_t] = g^{-1}\left(\beta_0 + f_1(x_{t1})\right)$$

where $f_p(x_{tp})$ is a spline basis, taking on the general form

$$f_p(x_{tp}) = \sum_{j=1}^{k} b_j(x)\beta_j \,. \tag{5}$$

Here, k is the number of basis vectors. This type of model called a **generalized additive model** (GAM). GAMs offer additional flexibility in modelling non-linear effects of predictors on the response $\{y_t\}$. By making the predictor a smooth function of the predictor variable, one can extend beyond mere linear combinations and describe more complex non-linear behaviours. The function $f(x_{tp})$ can be defined as a continuous polynomial or a piece-wise function. The case of a first order polynomial would reduce the GAM back to the form in Equation 3. While it would be possible to specify the degree of polynomial or piece-wise linear sections manually, it is also possible to let statistical software estimate the optimal[1] number of parameters.

Distributed lag and GLM Combining the distributed lag concept with the GAM results in what is typically called a **dynamic GAM** (dGAM). For a single covariate, it can be formulated as

$$E[Y_t] = g^{-1}\left(\beta_0 + \sum_{k=1}^{K} f(\mathbf{b_{k,t}}, x_{k,t})\right)$$

where $\mathbf{b_{k,t}}$ is a vector of κ number of coefficient defining the smooth for the functional effect of the data $x_{k,t}$ and K is the maximum number of lags. In most practical cases, where finite amounts of data is available, one would need to ensure that K is significantly smaller than the number of parameters in the model.

3.4 To Bayes, or not to Bayes

In the foreword of *Bayes Rules! An Introduction to Applied Bayesian Modeling*, Johnson, Ott, and Dogucu [21] write the following:

> "Though frequentist and Bayesian methods share a common goal – learning from data – the Bayesian approach to this goal is gaining popularity for many reasons: (1) Bayesian methods allow us to interpret new data in light of prior information, formally weaving both into a set of updated information; (2) relative to the confidence intervals and p-values utilized in frequentist analyses, Bayesian results are easier to interpret; (3) Bayesian methods can shine in settings where frequentist "likelihood" methods break down; and (4) the computational tools required for applying Bayesian techniques are increasingly accessible."

I will go through these claims one by one and discuss their relevance for our problem.

1. Bayesian inference asks what is plausible given limited data and a known structure, whereas frequentist inference asks what is supported by the data alone. That is, instead of merely finding optimal parameters based on maximum likelihood,

$$\underbrace{p(y \mid \theta)}_{\text{sampling distribution}} \tag{6}$$

   Bayesian inference seeks to find

$$\underbrace{p(\theta \mid y)}_{\text{posterior inference}} \propto \underbrace{p(y \mid \theta)}_{\text{sampling distribution}} \times \underbrace{p(\theta)}_{\text{structural information}} \tag{7}$$

   which is a distribution of model parameters θ given the observed data y. The structural information takes the form of a *prior distribution* over the parameter space. This distribution does

[1] The "optimal" number of parameters for a model can be estimated by minimizing a value such as the Akaike information criterion, a statistical measure which penalizes model complexity in order to avoid overfitting. The theory behind this is beyond the scope of this project. The interested reader is referred to section 3.1.4, 6.11 and 6.12 in Wood [18].

not depend on the data, but rather is informed by external knowledge. This means that physical laws, environmental boundaries or other expert knowledge can be used to restrict the possible parameter values. A simple example would be that of transport times. Even if we make no assumptions about the transport time from, say, Sellafield to Särdal, the least we can say is that it is not instant. There should be no effect of discharge at time t on activity concentrations at that same point in time. Such physical restrictions can be applied in the priors to constrain the model.

2. We are interested in the parameters that relate the discharge to the observed activity concentrations. Frequentist statistics gives us likelihood $p(y \mid \theta)$, which answers: *given this set of parameters θ, what is the probability that we observe this data y*? Bayesian inference provides a posterior $p(\theta \mid y)$, that answers: *given this set of data y, what is the probability that we observe these parameters θ?*. To link the former back to conclusions about the parameters requires some mental effort, whereas the latter Bayesian approach gives more direct insights, including credible intervals for estimated parameters. This is very useful for our particular research objective.

3. We have limited data available (see discussion in Section 4.3). Bayesian methods can be more effective in the case of low data availability, assuming appropriate priors are chosen [22].

4. The chosen software for data analysis and its use is discussed in Section 4.7

4 Approach

This section first discusses the characteristics of the data and the physics that will help define the model, as well as the challenges that may arise with that. Following that, the general principles of time series analysis are described. Finally, by building up from basic statistical principles, the distributed lag model is introduced and formulated for our particular problem and statistical toolkit.

4.1 Data description

The measurement data of isotopes studied in this project originate from Mattsson et al. [1] and are measurements of activity concentrations in Fucus *spp.* , reported in Bq per kg of dried weight of Fucus *spp.* (hereafter denoted Bq/kg d.w.), or else a unit derived from that quantity. All the measurement and discharge data in time-series form can be seen in Figure 6. The discharge data comes from various sources, and is provided on a yearly scale as TBq/y.

4.2 Physics considerations

The transport of radionuclides from nuclear facilities to the Swedish west coast, as well as consequent uptake by Fucus *spp.* are influenced by various physical processes:

1. **Process constraints:** The processing rate of spent fuel and radioactive waste is limited by factors like maximum throughput of chemical processes and staff capacity. The delivery of material is also a discrete, and not continuous process. Therefore, the amount of material processed – and consequently discharged – at a time t may be influenced by the amount processed at a time $t-1$. This can be quantified by the autocorrelation function (ACF), as shown in Figure 4.

2. **Accumulation:** Pollutants can accumulate in Fucus *spp.*, which like the first point also means there is some autocorrelation in the radioactivity measurements over time.

3. **Radioactivity:** Although accumulation occurs, there is also loss due to radioactivity following the isotope-specific half-life $T_{1/2}$. This diminishes activity concentrations over time.

4. **Biological loss:** The transfer of radionuclides out of Fucus *spp.* and back into the marine environment through various biological processes, as discussed in Section 2.2. This loss varies between isotopes, like how ^{129}I appears to accumulate in Fucus *spp.*, whereas other isotope activity concentrations like ^{137}Cs appear to diminish more rapidly much more quickly.

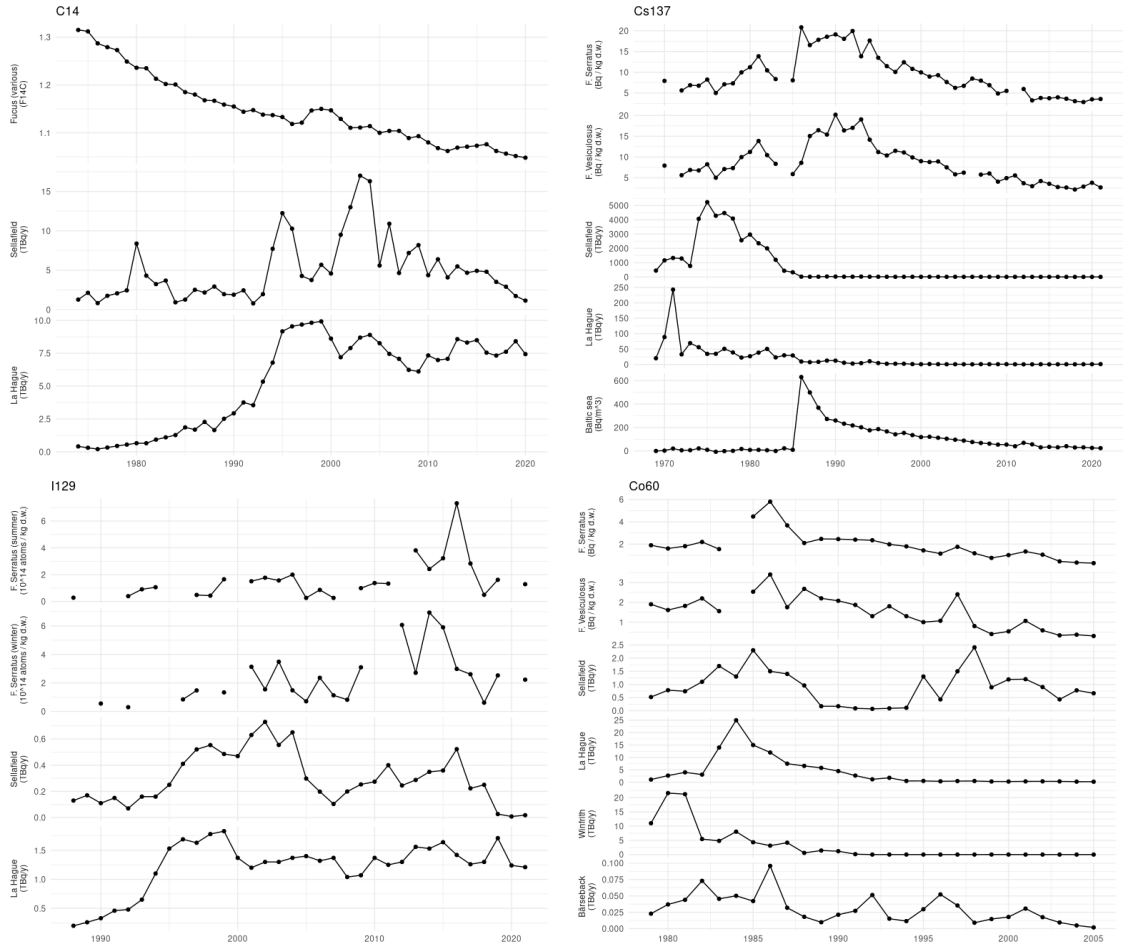


Figure 3: Time series plots, including all the predictors and measurements, for the four different isotopes: ^{14}C, ^{137}Cs, ^{129}I and ^{60}Co.

4.3 Dimensionality and data quality

It is also important to characterize the dataset itself, as there are some noteworthy features which affect the choice of method. The earliest and latest measurement, 1967-07-15 and 2023-12-31 respectively, were used to initialise a new dataframe with one `date` column, resulting in 20,682 observations (rows), which is 57 unique years. The available measurements from various isotopes were joined to this dataframe. This procedure was done mainly to make computational processing smoother. In reality only about 0.5% of the cells in the dataframe are filled with data. Table 1 shows the dates that have at least one measurement available. Because the discharge data is provided on a yearly basis, Table 1 also shows the number of years which have at least one measurement available.
Table 1 shows the limited amount of available data available per isotope. The non-trivial dynamics of radionuclide transport should encourage a parsimonious modelling approach. Over-parameterizing a model tends to lead to overfitting and also complicates interpretation. At the same time, it would be naïve not to inform our modelling approach at all by the physical and biological processes at hand. Therefore, the aim is to define a model that is flexible enough to account for the non-linear dynamical transport and uptake processes, while rigid enough to be interpretable. For the remainder of this study, the choice has been made to only look at four radionuclides with the most years of experimental data available, namely ^{14}C, ^{137}Cs, ^{60}Co and ^{129}I.

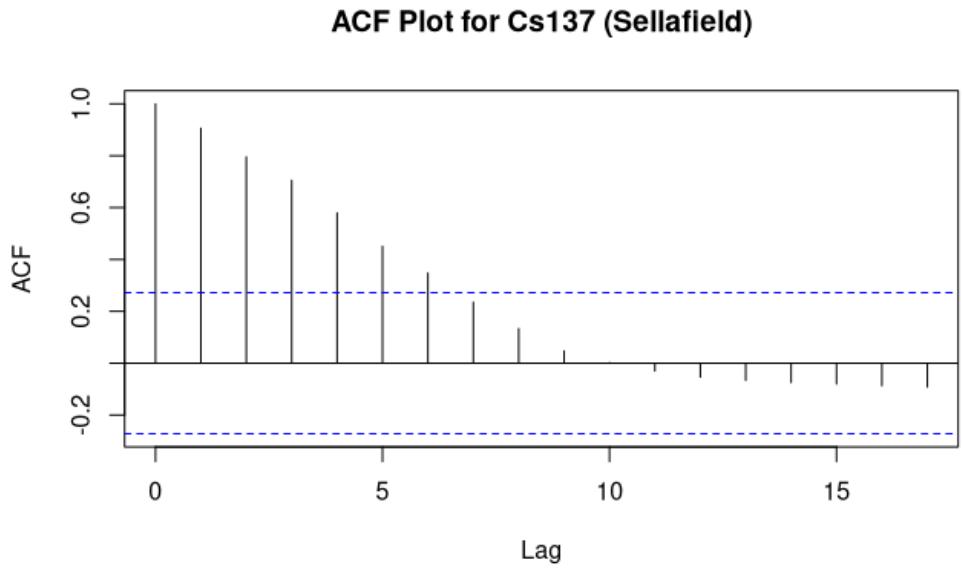


Figure 4: An example of the autocorrelation function (ACF) for various time lags of the yearly ^{137}Cs discharges from Sellafield. An ACF value outside of the region indicated by the dashed blue lines indicates a statistically significant correlation between the series at X_t and $X_{t-\tau}$. In this example, the previous 6 years are correlated with the current year's value. This is a statistical indication that the discharge is not random and the time series is non-stationary. The GAM model as described in Section 3.3 actually can deal with non-stationary data, and so does the software used in this project (see Section 4.7). ACF plots for all other data can be found in Appendix A.

4.4 Särdal measurement distributions

When constructing a model, it is important to consider how the response variable (in this case, the measurements in the seaweed for the four radionuclides) is distributed. Figure 5 shows the distribution of Särdal measurements in the four radionuclides, normal and log-transformed respectively. The latter shows that on the log scale, the measurement data is approximately normally (Gaussian) distributed.

4.5 Model terms

Now it is time to define the model terms. In the following paragraphs, we specify a model for one specific radioactive isotope and its corresponding discharges and sampled activity concentrations.

Background activity term We assume that there is some level of activity concentration in Fucus *spp.* even without the contribution of major external sources:

$$g(y_t) \propto \alpha + \varepsilon_t \tag{8}$$

where α is the mean activity concentration in Fucus *spp.* and $\varepsilon_t \sim \mathcal{N}(0, \sigma)$ is a random noise process accounting for small variations over time. This is intended to represent the natural background levels, which are assumed to be roughly constant. However, this term will also capture any 'residual' effect which is not included explicitly in the model, which could come from any discharge sources that are not explicitly added as predictors.

Autocorrelative term For Carbon-14 specifically, we should add an autocorrelative term to account for the bomb pulse:

$$g(y_t) \propto \alpha + u_t + \varepsilon_t \tag{9}$$

Särdal time series	Days with data	Years with data
C14 (mixed data)	72	54
Co60 (*F. serratus*)	127	29
Co60 (*F. vesiculosus*)	101	29
Cs134 (*F. serratus*)	81	19
Cs134 (*F. vesiculosus*)	28	11
Cs137 (*F. serratus*)	223	50
Cs137 (*F. vesiculosus*)	183	50
I129 (*F. serratus*)	81	37
Tc99 (*F. serratus*)	18	9
Tc99 (*F. vesiculosus*)	24	14

Table 1: For each radionuclide activity concentration measured from Särdal samples (the first column), the second column shows the number of days in the period $1967-2023$ for which there is was at least one measurement taken, and the third column shows the same on a yearly scale.

where $u_t = \rho u_{t-1}$ is an AR(1) term with ρ a parameter which will be estimated by the model. Note: it will be only used for capturing autocorrelation from the atmospheric CO_2 bomb pulse.

Species intercept Where sufficient seasonal (summer/winter) data is available, such information could be used to inform the model. There are two ways one might initially solve this. The first is *pooling*. This means creating a single model with one response series y_t. One simply aggregates the data into one time series, ignoring any potential variability between seasons. Alternatively, one could take a *no pooling* approach, using two separate response series and thus predict two separate models, e.g. y_t^{winter} and y_t^{summer} separately. This means no information is shared at all. Yet, there is a third approach. When one would like to account for possible variability among the species, yet have a monolithic model, a *partial pooling*[2] approach can be used. In practice, it means that the series are denoted by season using $seas$

$$g(y_t^{seas}) \propto \alpha + \alpha^{seas} + \varepsilon_t \tag{10}$$

where α^{seas} is a **season intercept**. That means, if there are two species, the model will include two intercepts, instead of one, yet these are not separate models. The parameters for the predictors (which are added in the next paragraph) will still be jointly estimated, hence this is only a *partial* pooling approach.

Distributed lag term Let the time series of marine discharges from the i-th source S^i, lagged by τ years, be denoted by $S^i_{t-\tau}$ where $\tau \geq 0$. The distributed lag term is then a summation of the smooth functions up to a maximum time lag L:

$$\sum_{\tau=0}^{L} f(S^i_{t-\tau}, \tau) \tag{11}$$

where

$$f(S_{t-\tau}, \tau) = \sum_{k=0}^{K} \sum_{m=0}^{M} \beta_{km} b_k(S_{t-\tau}) c_m(\tau) \tag{12}$$

is the bivariate tensor smooth, an extension of what was described in Section 3.3

General model The general model now becomes

[2]For an explanation of partial pooling, see Johnson, Ott, and Dogucu [21, Chapter 15].

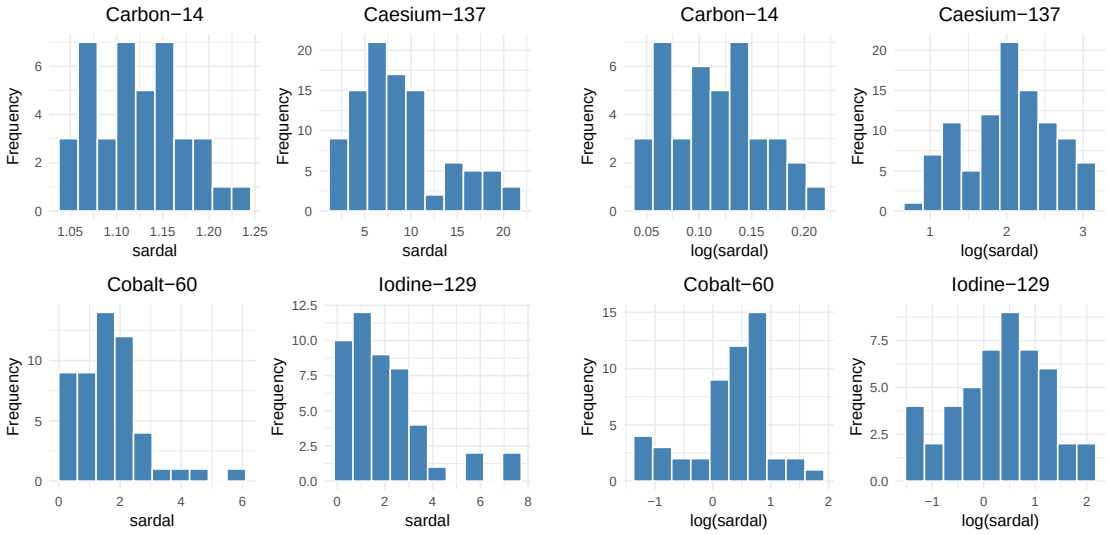


Figure 5: Histograms showing the distribution of Särdal measurements for the four selected isotopes. (a) The original measurements (Bq/kg d.w.), which are strictly positive and significantly right-skewed except for Carbon-14. (b) The same data on a log scale, where distributions are closer to normal. This supports modelling discharge measurements in Fucus *spp.* as $\log(y_t^{sp}) \sim \mathcal{N}(\mu_t, \sigma^2)$.

$$g(y_t^{\text{sp}}) = \underbrace{\alpha}_{\text{intercept}} + \underbrace{\alpha^{seas}}_{\text{seasonal effect}} + \sum_{i=0}^{j} \underbrace{\left(\sum_{l=0}^{L} f(\text{S}^i_{t-l}, l) \right)}_{\text{Contribution from source i}} + \underbrace{\varepsilon_t}_{\text{error term}} \tag{13}$$

where the AR(1) term $u_t = \rho u_{t-1}$, not included here, can be included if there is a significant trend in the measurements that is not explained by marine discharges.

Link function As discussed in Section 4.4, the measurements are approximately normally distributed on the log scale. Thus, we let $g(\cdot) = \log(\cdot)$, meaning that the contributions from the sources are additive on the log-linear scale, and multiplicative on the response scale.

4.6 Isotope-specific models

An isotope-specific model is specified for Carbon-14 (2 models), Caesium-137, Cobalt-60 and Iodine-129, leading to 5 separate dynamic GAM models following the general form described in Equation 13. The choice of distributed lag terms is based on what are considered to be likely dominant contributing sources while balancing with data availability. This has been motivated in Section 2.2 and 4.3.

Carbon-14 There is a major contribution from the atmospheric bomb pulse. We try two approaches. First, simply adding an AR(1) component to try and capture the residual bomb pulse effect. With Sellafield (SF) and La Hague (LH) as sources, the model, written explicitly, becomes

$$\log(y_t) = \alpha + \sum_{l=0}^{L} f(\text{SF}_{t-l}, l) + \sum_{l=0}^{L} f(\text{LH}_{t-l}, l) + u_t + \varepsilon_t \tag{14}$$

where $u_t = \rho u_{t-1}$ is the AR(1) term. The prior set for the AR term is $\rho \sim \text{Beta}(2, 2)$, as this prior is bounded on the interval $[0, 1]$, with the majority of the weight of this distribution towards the middle of that interval.

The second approach is to modify y_t instead, subtracting a scaled version of the atmospheric bomb pulse to approximate a time series without the bomb pulse. The data for atmospheric $F^{14}C$ between 1950-2019 is taken from Hua et al. [23], with linear extrapolation afterwards to match the dimensionality of the Särdal measurements. Although in reality the bomb pulse in the marine environment is more complicated, depending on both the specific site and the depth, this approximation was used to de-trend the time series in order to see if this increased the potential for the model to pick up on influences from SF and LH. The modified time series is $y_t^* = y_t - 0.68 \cdot y_t^{bomb}$. With SF and LH, the model now becomes

$$\log(y_t^*) = \alpha + \sum_{l=0}^{L} f(\mathrm{SF}_{t-l}, l) + \sum_{l=0}^{L} f(\mathrm{LH}_{t-l}, l) + \varepsilon_t \tag{15}$$

Caesium-137 Sellafield is considered a likely major influences, but as discussed in Section 2.2, the primary contribution is considered Baltic sea outflow containing ^{137}Cs from the Chernobyl disaster. A synthetic Baltic sea outflow (BS) dataset was generated using the estimated fallout quantities and effective half lives described in Herrmann et al. [11] [3]. The model, written explicitly, becomes

$$\log(y_t) = \alpha + \sum_{l=0}^{L} f(\mathrm{SF}_{t-l}, l) \sum_{l=0}^{L} f(\mathrm{BS}_{t-l}, l) + \varepsilon_t \tag{16}$$

Iodine-129 The data provided had only samples for F. Serratus, but the ^{129}I dataset is unique in having sufficient sampling done by season. As discussed in Section 2.2, ^{129}I activity concentrations in F. Serratus have been observed to significantly differ between summer and winter samples. Thus, we will model the summer and winter samples in the same model, but allow for a season-specific intercept. This can capture variation in levels between the seasons. As for external influences, we include Sellafield and La Hague:

$$\log(y_t^{seas}) = \alpha + \alpha^{seas} + \sum_{l=0}^{L} f(\mathrm{SF}_{t-l}, l) + \sum_{l=0}^{L} f(\mathrm{LH}_{t-l}, l) + \varepsilon_t z \tag{17}$$

where *seas* now denotes the season, being winter or summer.

Cobalt-60 The influences modelled are La Hague (LH) and Barsebäck (BB). Due to low data availability, the model cannot be made overly complex. Therefore, only LH and BB are included and Winfrith, although possibly significant for early in the time series, is not included. Written explicitly, the model becomes

$$\log(y_t) = \alpha + \sum_{l=0}^{L} f(\mathrm{LH}_{t-l}, l) + \sum_{l=0}^{L} f(\mathrm{BB}_{t-l}, l) + \varepsilon_t \tag{18}$$

4.7 Model specification in `mvgam`

Ecological time series often consist of discrete observations and a multivariate model consisting of environmental predictors. The R package `mvgam` was introduced by Clark and Wells [24], with the goal of making Bayesian specification of dynamic GAMs more straightforward, while addressing some of the common problems[4] that are present in ecological time series analysis, such as observation errors, missing values and/or irregular sampling (due to limited capacity or sample loss). Often, the data is also overdispersed, which means the observed data has larger variance than can be explained by the parametric model. Because `mvgam` is designed to make the model specification smoother for such modelling problems, and the fact that GAM models naturally handle non-stationary data (which the

[3] For the R code used to generate this data, see Appendix B.

[4] See the introduction of Clark and Wells [24] and the references therein for more detailed explanation of common ecological time series challenges. Here, we have just summarized the main challenges relevant for the radionuclide dataset.

data in this project is; see Section 4.2) it is the modelling tool of choice for this project.

We used `mvgam` (version 1.1.593; Clark and Wells [24]) to construct, fit and interrogate the model. `mvgam` incorporates functionalities from the `brms` (Bürkner [25]), `mgcv` (Wood [18]) and `splines2` (Wang and Yan [20]) packages. The mvgam-constructed model and observed data were passed to the probabilistic programming environment Stan (version 2.37.0; Carpenter et al. [26], Stan Development Team 2026), specifically through the `cmdstanr` interface (Gabry et al. [27]). For all isotope models, 4 Hamiltonian Monte Carlo chains were ran for 2000 warmup iterations and 4000 sampling iterations for joint posterior estimation. Rank normalized split Rhat ([28]) and effective sample sizes were used to monitor convergence. Because of low data availability, the lag and discharge spline bases were specified at a complexity $K = 4$, which allows non-linear response while restricting the number of parameters.

Because of the low data availability, the decision was made not to leave out any data for checking inference of the model. It is important to note that this means we cannot truly say whether the model is overfitting or not. However, there are some indications that can be used:

1. The number of effective degrees of freedom (EDFs) is a measure of the effective number of parameters needed to describe the fitted model. Because `mvgam` penalizes complexity in the model via regularization methods, a total number of EDFs that is significantly lower than what was allowed indicates that the software favours low-complexity model over a high-complexity one. By definition, a model that fits to noise is one of higher complexity than one that does not fit to noise.

2. The actual fitted timeseries, back-casted onto the data. If we see the fitted time series perfectly fitting to each data point, even the outliers, this suggests overfitting.

3. Autocorrelation in the residuals. This is actually a measure for *underfitting* (which we will use it for), but a total absence of autocorrelation in the residuals can signify that the model is *overfitting*, as it captures all the noise in the model.

5 Results

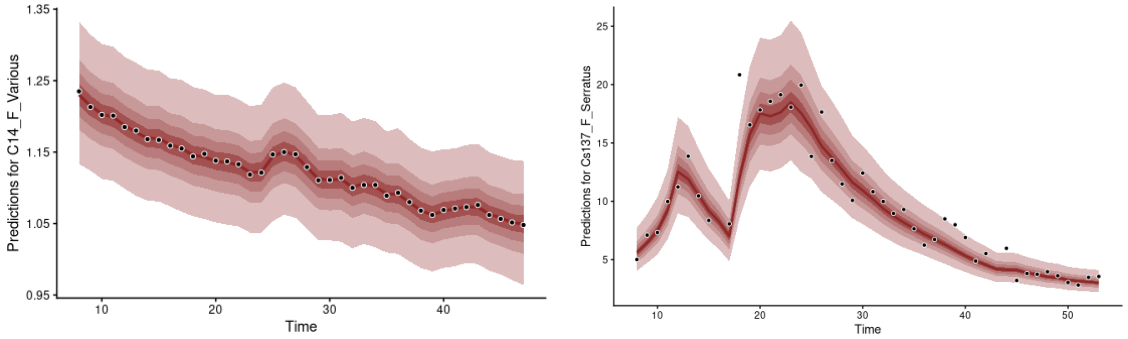
5.1 Fitted timeseries

The number of iterations specified in Section 4.7 resulted in Rhat of 1 for all chains of all models. This indicates a convergence and that `mvgam` was able to find a fit for the data. The fitted time series models are displayed in Figure 6. Overall the models appear to capture the trends well. However, it is important to diagnose the model further. This is done in the next section.

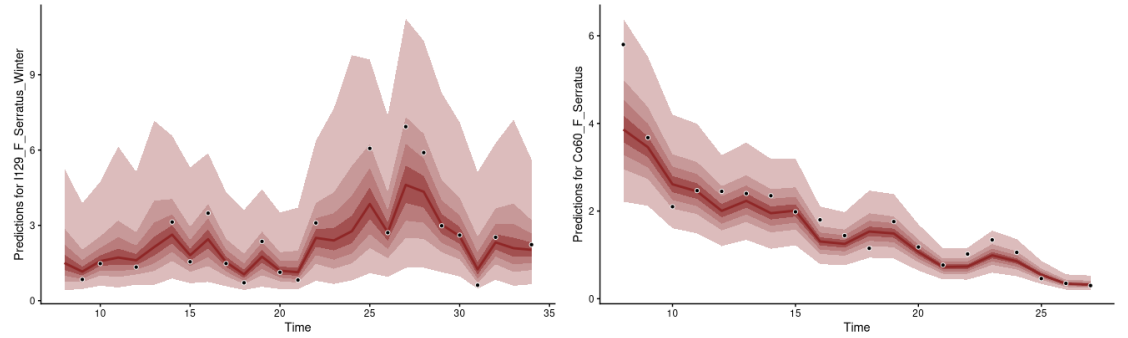
5.2 Model diagnostics

We check three things: first, the effective degrees of freedom (EDFs). We compute the ratio of EDFs to the maximum allowed degrees of freedom to see if `mvgam` favoured a less complex model over a more complex model. Secondly, we will check whether there is any autocorrelation in the residuals. That would indicate that there is some effect that is not captured by the model. Lastly, we will perform a posterior predictive check (PPC), which is a common way to check to check the overall fit of a Bayesian model [29].

Effective degrees of freedom See Table 2 for a summary of the EDF statistics of the fitted models. A lower ratio would imply a simpler model, meaning `mvgam` has regularized the model effectively. This suggests a correctly fitted model. On the other hand, a ratio closer to 1 means the signal in the data is likely very weak, that is, there is no clear influence from the discharges on the observed concentrations under the specified model. Like mentioned before, this *can* indicate overfitting, but is not a guaranteed indicator for that.



(a) Carbon-14 (AR-based model) **(b) Caesium-137 (F. Serratus)**



(c) Iodine-129 (F. Serratus, Winter) **(d) Cobalt-60 (F. Serratus)**

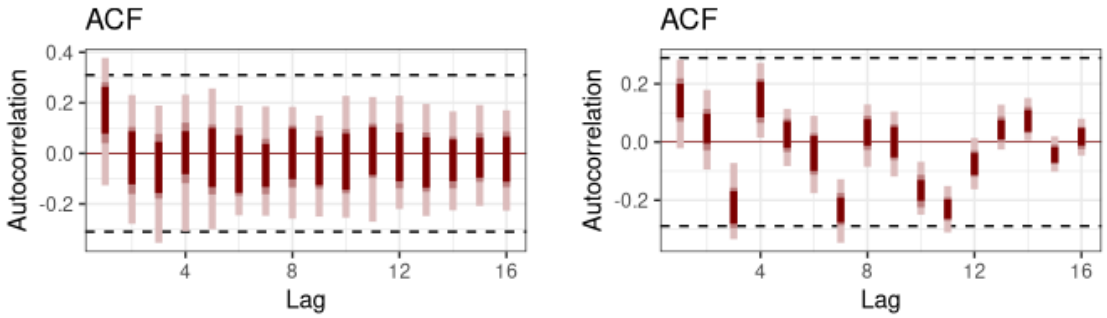
Figure 6: Fitter time series for each isotope. Results are only shown for select time series. The full results are in Appendix C.

Isotope	Max. DoFs	EDFs	Complexity
^{14}C (AR-based model)	30	20.91	0.70
^{137}Cs	46	4.893	0.11
^{129}I	32	11.51	0.36
^{60}Co	20	5.035	0.25

Table 2: Maximum allowed degrees of freedom (DoFs) compared to the effective degrees of freedom (EDFs). The last column, complexity, is defined here as the ratio of EDFs to Max. DoFs. A value close to 1 means the model took 'advantage' of allowed complexity, a lower ratio means it has regularized strongly.

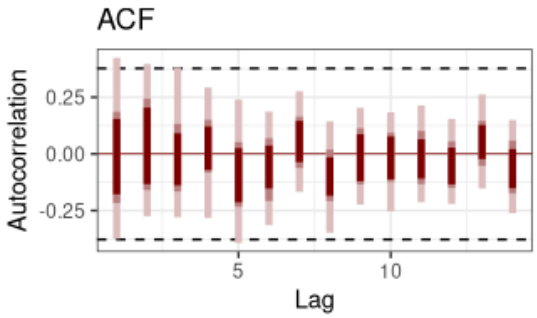
Residual autocorrelation If there is any significant autocorrelation left in the residuals $y - \hat{y}$ of the model, then this implies there is some discharge or other effect causing activity concentrations to change, that is not one of the predictors included in the model. On the other hand, no significant autocorrelation, although it once again does not confirm this is the *best* model, tells that there is at least no significant effect left uncaptured under the given fitted model. For all models, the discharges appear to explain most of the observed concentration. Winfrith, which was left out of the ^{60}Co model due to low data availability, thus does not appear a major influence, otherwise significant autocorrelation would be left in the ^{60}Co series.

Posterior predictive checks The idea behind a PPC is to see whether a fitted model can replicate data that looks similar to the original distribution [29, 30]. This is not necessarily an indication of the *best* fit, since multiple (possibly infinitely many) different statistical models could fit the same data and produce representative results [30]. It is however a straightforward way to see if the model can reproduce time series similar to the observed data. Figure 8 shows the PPC plots for each isotope model. Overall the models can reproduce representative data but fail to capture the extreme outliers (i.e. sharp peaks in measured activity concentration).

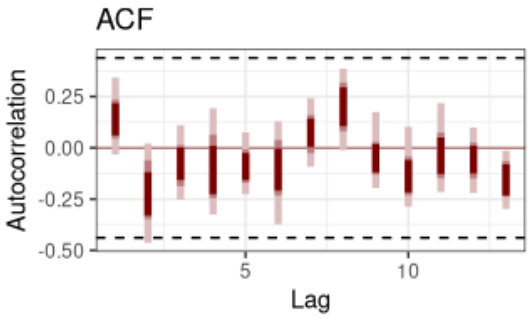


(a) Carbon-14 (AR-based model)

(b) Caesium-137



(c) Iodine-129



(d) Cobalt-60

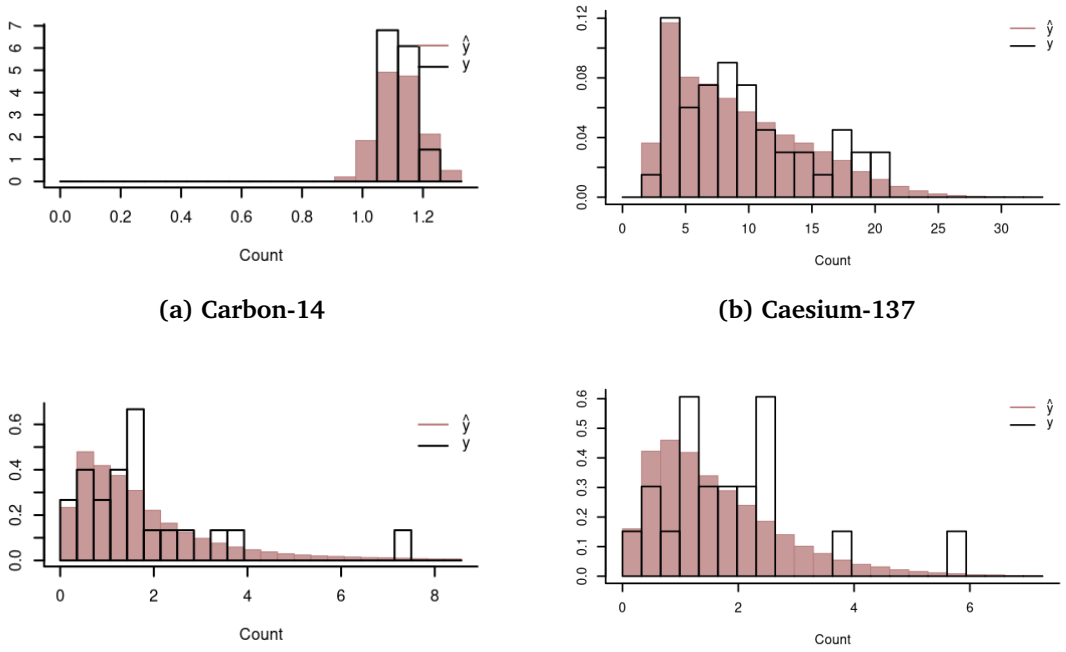
Figure 7: Auto-correlation functions for the residuals of each isotope model. A dark red line outside of the dashed black lines would indicate statistically significant autocorrelation. Only lag 7 of the ^{137}Cs model appears to have some significant residual autocorrelation. Overall, the models are able to describe Särdal activity concentrations as a function of the marine discharges without any major influence still left. Carbon-14 shows very little deviation from the zero-line, which, combined with the high complexity factor in Table 2, may indicate an overfitted model. For the other isotopes, that there is some variation in the distribution of the residuals, which at least indirectly indicates that these models are not fitting to *all* the noise, indicating neither a significant overfit nor a significant underfit.

5.3 Discharge-lag fitted surfaces

Figure 9-12 show the actual fitted tensor smooth surfaces, showing the joint effect of discharge and lag on the observed activity concentrations in Fucus *spp.* at Särdal. A common observation between the ^{137}Cs, ^{129}I and ^{60}Co models is the very broad, dispersed structure in the surface. This suggests that the discharges do not just have an effect on measured activity concentrations upon initial arrival, but rather, there is a strongly diffusive effect that spreads the uptake and consequent observation over several years. This kind of structure is one of the key insights that a distributed-lag approach is able to give.

5.4 Additional terms in ^{14}C and ^{129}I

The measurements in ^{14}C most clearly showed a trend, which comes from the atmospheric bomb pulse. We accounted for this in two ways, one of which was using an AR(1) component. ^{129}I was the only isotope with sufficient and seasonally balanced data for the dGAM model to have a seasonal effect included. Figure 13 shows the results for these two special terms.



(a) Carbon-14 **(b) Caesium-137**

(c) Iodine-129 **(d) Cobalt-60**

Figure 8: PPC plots for each isotope on the original response scale ($F^{14}C$ for ^{14}C, Bq / kg d.w. for the others). Here, y (black) represents the true distribution of the data, and $\hat{y}$ represents the posterior draws. Overall we can see that the model can reproduce representative data, although the model fails to capture the extreme values (data far to the right) of the data.

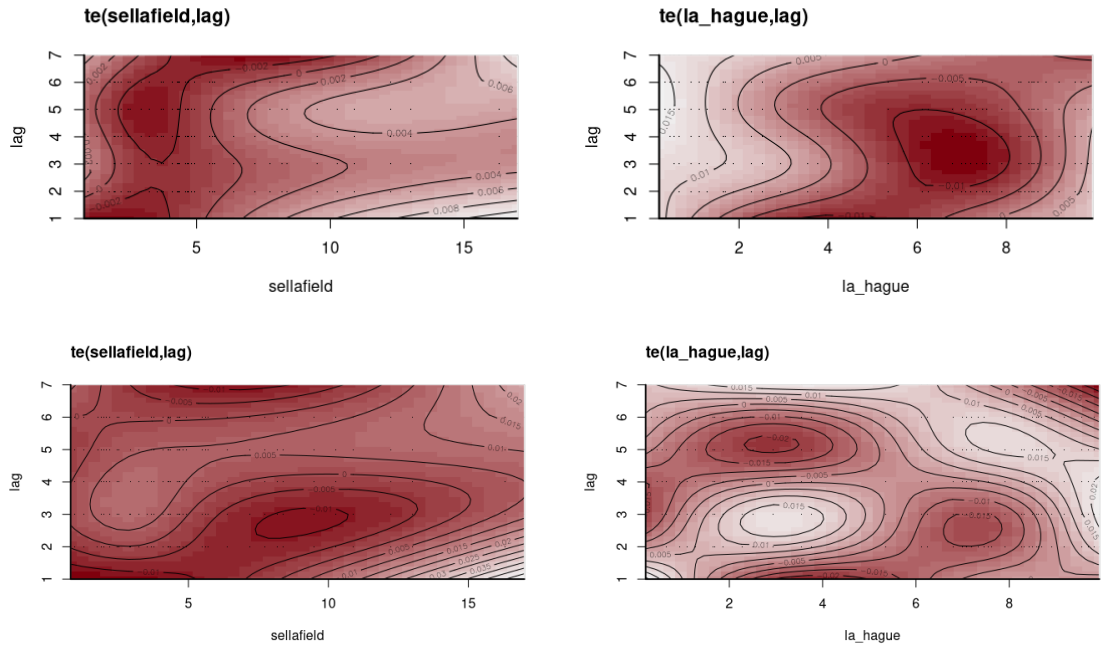


Figure 9: Fitted tensor smooth surfaces and time series for ^{14}C, for the AR(1) model (top row) and the bomb-pulse corrected model (bottom row). For both fits, the overall shape in both dimensions and for both sources is not aligning with previous literature. Keeping in mind that the previous diagnostics indicating possible poor fit, this seems to further corroborate that the model is overfitting (over complex and fitting to noise).

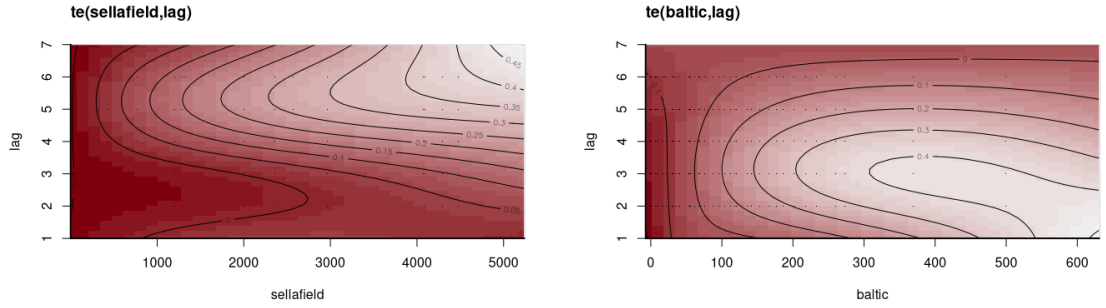
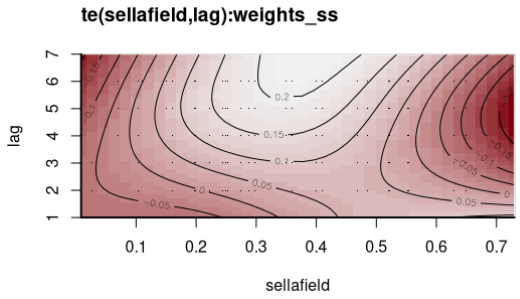
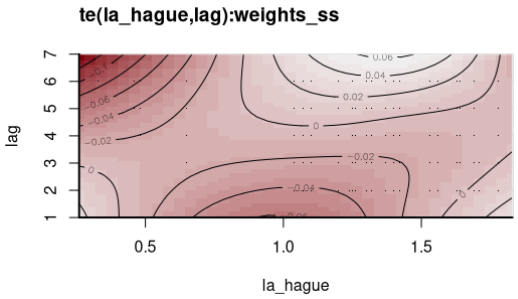


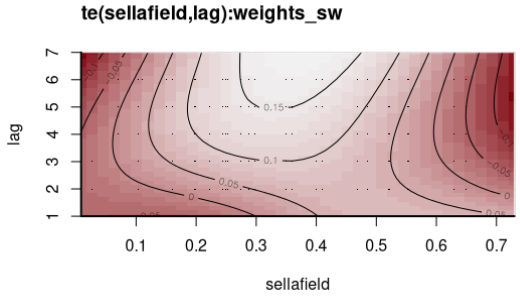
Figure 10: Fitted tensor smooth surfaces and time series for ^{137}Cs. The effect of Baltic sea outflow is concentrated around 1-2 years for maximum Baltic activity concentrations, but with a broad effect across lags. This agrees with expectation of transport time into Kattegat. Sellafield discharges show an increasing effect with increasing absolute discharge, with initial change in surface gradient around 4-5 years, which aligns reasonably with literature estimated transport time of 3-4 years from this source [1]. Furthermore, a broad effect, possibly beyond 7 years (although this is not modelled due to limited data), is observed.



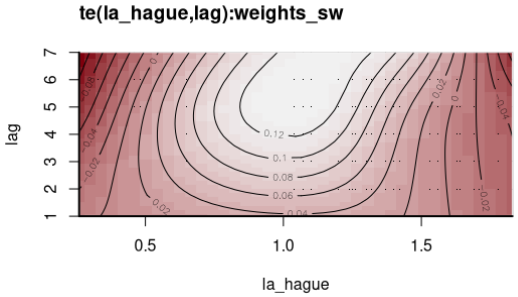
(a) Sellafield (summer)



(b) La Hague (summer)

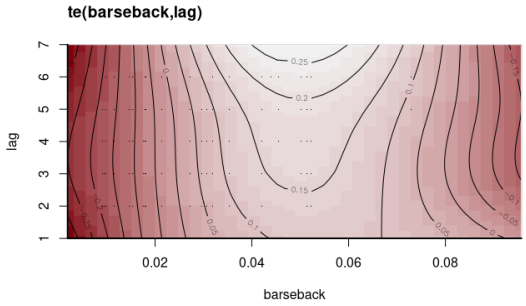


(c) Sellafield (winter)



(d) La Hague (winter)

Figure 11: Fitted tensor smooth surfaces and time series for ^{129}I, by season. The overall shape in the lag dimension is aligning well with expectations in winter samples, with the strongest effect starting around 4-5 years and 3-4 years for Sellafield and La Hague respectively.



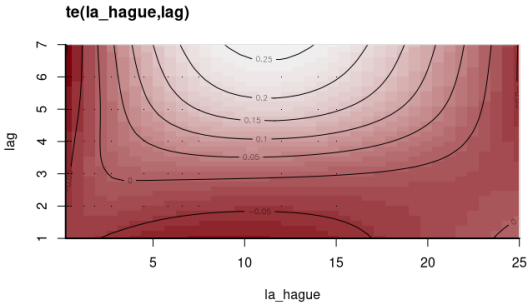


Figure 12: Fitted tensor smooth surfaces and time series for ^{60}Co. Barsebäck shows a broad effect across all time lags, which would indicate a sustained (albeit low) effect in the marine environment. La Hague has a more defined gradient over the lag dimension, particularly increasing around 4-5 years after the discharge, which is slightly more delayed than what was found for ^{129}I in Figure 11.

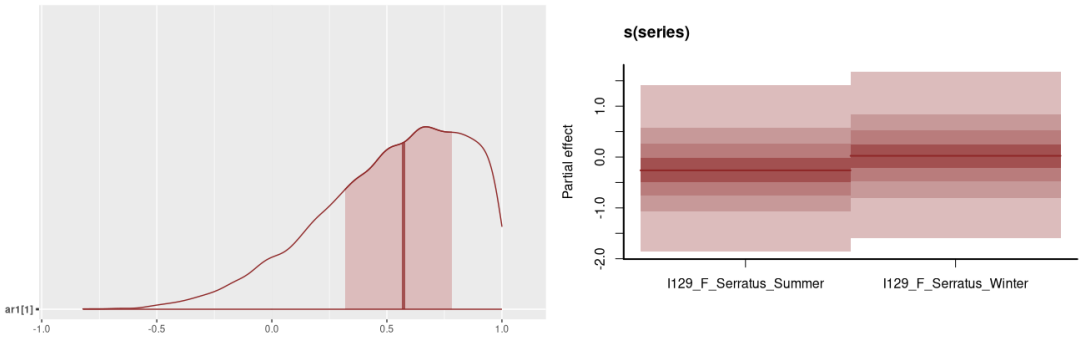


Figure 13: The estimated autoregression parameter estimate for ^{14}C (left) and the partial seasonal effects for ^{129}I (right). The mean of 0.57 for the parameter ρ in Equation 14 shows a significant effect of the bomb pulse, as expected. The majority of the weight of the estimated distribution lies towards 1, which indicates that the bomb pulse dominates the observed concentration, as expected. This further highlights the difficulty in this particular model, as almost all of the signal is captured in the AR(1) term. For the seasonal effect in ^{129}I, there appears to be a marginal difference in the estimated summer and winter effect, although the uncertainty on these estimated seasonal effects is very large.

6 Conclusion

The dynamic generalized additive model proposed in this project has carefully been built up from first principles of time analysis, with the concept of lag-discharge tensor smooth surfaces used as a way to describe the non-linear transport process. While some of the models managed to reproduce some results that are in line with literature, and give some new insight into the non-linear relationship between discharge and observed levels in Fucus *spp.*, it is difficult to say whether the models are truly representative or not. A more rigorous approach such as leave-one-out cross-validation (LOOCV) or some type of sensitivity analysis is required to thoroughly test the models, but within the scope of this project there was not enough time.

Rather, the focus of this project was to explore whether dGAM modelling approach is suitable at all for radionuclide transport in the marine environment. The flexibility of the dGAM approach is clear, although perhaps is a bit *too* flexible. By the way the tensor smooth surfaces are constructed and fitted, it is not possible impose physical restrictions upon the shape of the surface in any dimension, lag or discharge. However, there are reasonable restrictions that could be imposed on a radionuclide transport model. The most poignant example of this is the impossibility of instantaneous or near-instantaneous effect of remote marine discharges on Särdal levels. Take for example Sellafield discharges. Without assuming the exact transport time, we expect a negligible effect at the lag 0, 1 and 2. Therefore, it would be beneficial to penalize these improbable or impossible lags. However, such a restriction is not possible to impose in a dGAM. One way is to exclude lags that are unlikely to have an effect from the data completely. While rather crude, it is justifiable to exclude at least lag 0 (instant transport) from the data, which is what was done in the approach of this project.

Overall, the conclusion is that the distributed lag modelling has potential for application to radionuclide transport, as the models for ^{129}I and ^{137}Cs showed reasonable agreement with literature. However, there are two primary limiting factors for the effectiveness of the model. First of all, having bivariate smooths in both lag and discharge dimensions adds significant complexity. The fact that the formalism of dGAMs does not allow for restrictions on the actual shape of the tensor smooth surface, makes it difficult to reliably apply it in its current form. A second limiting factor is the low data availability combined with lack of some key data (such as lack of data on ^{14}C releases from Ringhals NPP and Barsebäck NPP).

7 Future Outlook

Inclusion of Ringhals NPP and Barsebäck NPP The geographic proximity of Ringhals NPP and Barsebäck NPP to Särdal means that these sources of effluent containing radionuclides should be accounted for. This was done for ^{60}Co but not the other ones. As discussed before, ^{14}C release into the marine environment is not monitored at these sites, meaning the influence on the observed concentrations in Fucus *spp.* cannot be quantified using this kind of model. Future models should attempt to account for the releases from these two nuclear sites, where such data is available.

Restrictive modelling While the lag-discharge tensor smooth surfaces in theory provide a lot of flexibility, as discussed in the conclusion, there are actually reasonable priors one could impose on the relation. Future work could include finding alternative distributed-lag-type models where physics restrictions can explicitly be used to penalize unrealistic influence from early lags.

Mixed temporal scales In the current approach and method, the temporal scales of predictors and the response have to be matched in order to create a functional model in `mvgam`. When one has two different temporal scales (daily measurements and yearly discharges), there are two solutions: 1) upsampling the yearly discharges to daily discharges (for example assuming constant discharge throughout the year); or 2) downsampling the daily measurements to yearly scale (e.g. averaging). The second approach was chosen in this project. It is clear that either procedure leads to a loss of information. Future work could thus include 1) sourcing higher frequency discharge data, e.g. monthly or daily discharges from Sellafield, La Hague, et cetera; or 2) explore mixed-frequency methods. This

approach was developed in econometrics and is used in cases where one wants to model e.g. quarterly GDP using daily stock market prices as a predictor [31]. Either of these would increase the total amount of information available in the model and thus in theory improve chances of finding a suitable model, but were outside the scope of this project.

Dilution factors The dGAM model can flexibly deal with various scales and units in both the predictors and the response variable, where the model presented here uses absolute discharge quantities in TBq as a predictor for activity concentration in seaweed Bq/kg d w. This works fine for obtaining information about lags and thus transport times. However, in order to obtain dilution factors directly from the presented model, the predictors and response would need to be matched in units. Specifically, one would need time series of measurements of seaweed (ideally Fucus *spp.*) very close to the discharge pipe exits at both La Hague and Sellafield (and possibly other sources) so that an activity concentration in $AC_{\text{source}} \sim$ Bq/kg d w can be obtained. With such data as predictors, instead of the current predictors in TBq, the parameters in a model $AC_{\text{seaweed}} \sim \sum_i \beta_i AC_{\text{source},i}$ become interpretable as dimensionless dilution factors. However, linking such data is further complicated by different effects of temperature and salinity of the local marine environment. As such, it is unlikely that the presented models could be used in a meaningful way to determine dilution factors.

Seasonality The model appears to pick up on seasonal effect for the ^{129}I data, supporting recent findings by Mattsson et al. [1]. This result highlights that some kind of hierarchical modelling approach, with seasons as levels, could effective for investigating seasonal effects for other isotopes too. However, allowing for variation in the season explicitly leads to increased model complexity, since this adds parameters to the model. Such an approach can thus only really be taken with sufficient (season-specific) data available for a given isotope.

References

[1] Sören Mattsson et al. “Radionuclides in Algae from Swedish Coastal Waters for over Half a Century”. In: *Radiation Protection Dosimetry* 201.13–14 (Aug. 26, 2025), pp. 974–994. ISSN: 0144-8420, 1742-3406. DOI: 10.1093/rpd/ncaf070.

[2] Kristina Eriksson Stenström et al. *Carbon-14 in the Marine Environment of Ringhals Nuclear Power Plant*. Report for project SSM2022-4035. ISSN: 2000-0456. 2025:08. Swedish Radiation Safety Authority (SSM), Aug. 2025.

[3] P. Lindahl et al. “Long-Term Study of 99Tc in the Marine Environment on the Swedish West Coast”. In: *Journal of Environmental Radioactivity* 67.2 (2003), pp. 145–156. ISSN: 0265931X. DOI: 10.1016/S0265-931X(02)00176-5.

[4] David McCubbin et al. “Distribution of Technetium-99 in Sub-Tidal Sediments of the Irish Sea”. In: *Continental Shelf Research* 26.4 (Mar. 2006), pp. 458–473. ISSN: 02784343. DOI: 10.1016/j.csr.2005.12.010.

[5] Kristina Eriksson Stenström and Sören Mattsson. “Spatial and Temporal Variations of 14C in Fucus Spp. in Swedish Coastal Waters”. In: *Journal of Environmental Radioactivity* 242 (Feb. 2022), p. 106794. ISSN: 0265931X. DOI: 10.1016/j.jenvrad.2021.106794.

[6] Finansdepartementet. *Long-Term Investments in Nuclear Power for Swedish Electricity Supply*. Press release. Finansdepartementet (Ministry of Finance in Sweden), Oct. 28, 2025.

[7] V I Khripunov, D K Kurbatov, and M L Subbotin. “C-14 Production in CTR Materials and Blankets”. In: ().

[8] UNSCEAR. *Sources, Effects and Risks of Ionizing Radiation*. UNSCEAR 2008 Report. 2008, Volume 1. Annex B. Exposures of the public and workers from various sources of radiation.

[9] Kieran Michael Tierney. “Marine Ecosystem Uptake of Nuclear Reprocessing Derived Radiocarbon (14C)”. Glasgow: University of Glasgow, 2017.

[10] Agata Zaborska, Aleksandra Winogradow, and Janusz Pempkowiak. “Caesium-137 Distribution, Inventories and Accumulation History in the Baltic Sea Sediments”. In: *Journal of Environmental Radioactivity* 127 (Jan. 1, 2014), pp. 11–25. ISSN: 0265-931X. DOI: 10.1016/j.jenvrad.2013.09.003.

[11] J. Herrmann et al. “Radioactivity in the Baltic Sea, 1999-2006 HELCOM Thematic Assessment”. In: (2009).

[12] X. L. Hou et al. “Level and Origin of Iodine-129 in the Baltic Sea”. In: *Journal of Environmental Radioactivity* 61.3 (Jan. 1, 2002), pp. 331–343. ISSN: 0265-931X. DOI: 10.1016/S0265-931X(01)00143-6.

[13] G. T. Cook et al. “Remobilization of Sellafield-derived Radionuclides and Transport from the North-East Irish Sea”. In: *Journal of Environmental Radioactivity* 35.3 (Jan. 1, 1997), pp. 227–241. ISSN: 0265-931X. DOI: 10.1016/S0265-931X(96)00070-7.

[14] Yuhi Satoh, Shigeki Wada, and Shun’ichi Hisamatsu. “Seasonal Variations in Iodine Concentrations in a Brown Alga (Ecklonia Cava Kjellman) and a Seagrass (Zostera Marina L.) in the Northwestern Pacific Coast of Central Japan”. In: *J Oceanogr* 75.1 (Feb. 1, 2019), pp. 111–117. ISSN: 1573-868X. DOI: 10.1007/s10872-018-0479-8.

[15] Udo Nitschke et al. “Variability in Iodine in Temperate Seaweeds and Iodine Accumulation Kinetics of *Fucus Vesiculosus* and *Laminaria Digitata* (Phaeophyceae, Ochrophyta)”. In: *Journal of Phycology* 54.1 (Feb. 2018). Ed. by C. Hurd, pp. 114–125. ISSN: 0022-3646, 1529-8817. DOI: 10.1111/jpy.12606.

[16] Andrew B. Cundy et al. “Decline of Radionuclides in the Nearshore Environment Following Nuclear Reactor Closure: A U.K. Case Study”. In: *Environ. Sci. Technol.* 33.17 (Sept. 1, 1999), pp. 2841–2849. ISSN: 0013-936X. DOI: 10.1021/es9811694.

[17] Robert H. Shumway and David S. Stoffer. “Time Series Regression and Exploratory Data Analysis”. In: *Time Series Analysis and Its Applications*. Cham: Springer Nature Switzerland, 2025, pp. 49–83. ISBN: 978-3-031-70583-0 978-3-031-70584-7. DOI: 10.1007/978-3-031-70584-7_2.

[18] Simon N. Wood. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC, May 18, 2017. ISBN: 978-1-315-37027-9. DOI: 10.1201/9781315370279.

[19] Shirley Almon. “The Distributed Lag Between Capital Appropriations and Expenditures”. In: *Econometrica* 33.1 (Jan. 1965), p. 178. ISSN: 00129682. DOI: 10.2307/1911894. JSTOR: 1911894.

[20] Wenjie Wang and Jun Yan. “Shape-Restricted Regression Splines with R Package Splines2”. In: *Journal of Data Science* 19.3 (Aug. 12, 2021), pp. 498–517. ISSN: 1680-743X, 1683-8602. DOI: 10.6339/21-JDS1020.

[21] A. Johnson, Q. Ott, and Mine Dogucu. *Bayes Rules! An Introduction to Applied Bayesian Modeling*. Chapman & Hall/CRC.

[22] Daniel McNeish. "On Using Bayesian Methods to Address Small Sample Problems". In: *Structural Equation Modeling: A Multidisciplinary Journal* 23.5 (Sept. 2, 2016), pp. 750–773. ISSN: 1070-5511. DOI: 10.1080/10705511.2016.1186549.

[23] Quan Hua et al. "ATMOSPHERIC RADIOCARBON FOR THE PERIOD 1950–2019". In: *Radiocarbon* 64.4 (Aug. 2022), pp. 723–745. ISSN: 0033-8222, 1945-5755. DOI: 10.1017/RDC.2021.95.

[24] Nicholas J. Clark and Konstans Wells. "Dynamic Generalised Additive Models (DGAMs) for Forecasting Discrete Ecological Time Series". In: *Methods Ecol Evol* 14.3 (Mar. 2023), pp. 771–784. ISSN: 2041-210X, 2041-210X. DOI: 10.1111/2041-210X.13974.

[25] Paul-Christian Bürkner. "Brms: An R Package for Bayesian Multilevel Models Using Stan". In: *Journal of Statistical Software* 80 (Aug. 29, 2017), pp. 1–28. ISSN: 1548-7660. DOI: 10.18637/jss.v080.i01.

[26] Bob Carpenter et al. "Stan: A Probabilistic Programming Language". In: *J Stat Softw* 76 (2017), p. 1. ISSN: 1548-7660. DOI: 10.18637/jss.v076.i01. PMID: 36568334.

[27] Jonah Gabry et al. *Cmdstanr: R Interface to 'CmdStan'*. manual. 2025.

[28] Aki Vehtari et al. "Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion)". In: *Bayesian Analysis* 16.2 (June 2021), pp. 667–718. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/20-BA1221.

[29] Nicholas J. Clark. *Posterior Predictive Checks for Mvgam Models*. Reference Manual.

[30] Andrew Gelman et al. *Bayesian Data Analysis*.

[31] Han Lui and Ying Lui. "Mixed-Frequency Models". In: *Econometric Modelling and Forecasting of Tourism Demand*. Routledge, Oct. 27, 2022, pp. 144–172. DOI: 10.4324/9781003269366-7.

A Autocorrelation tests

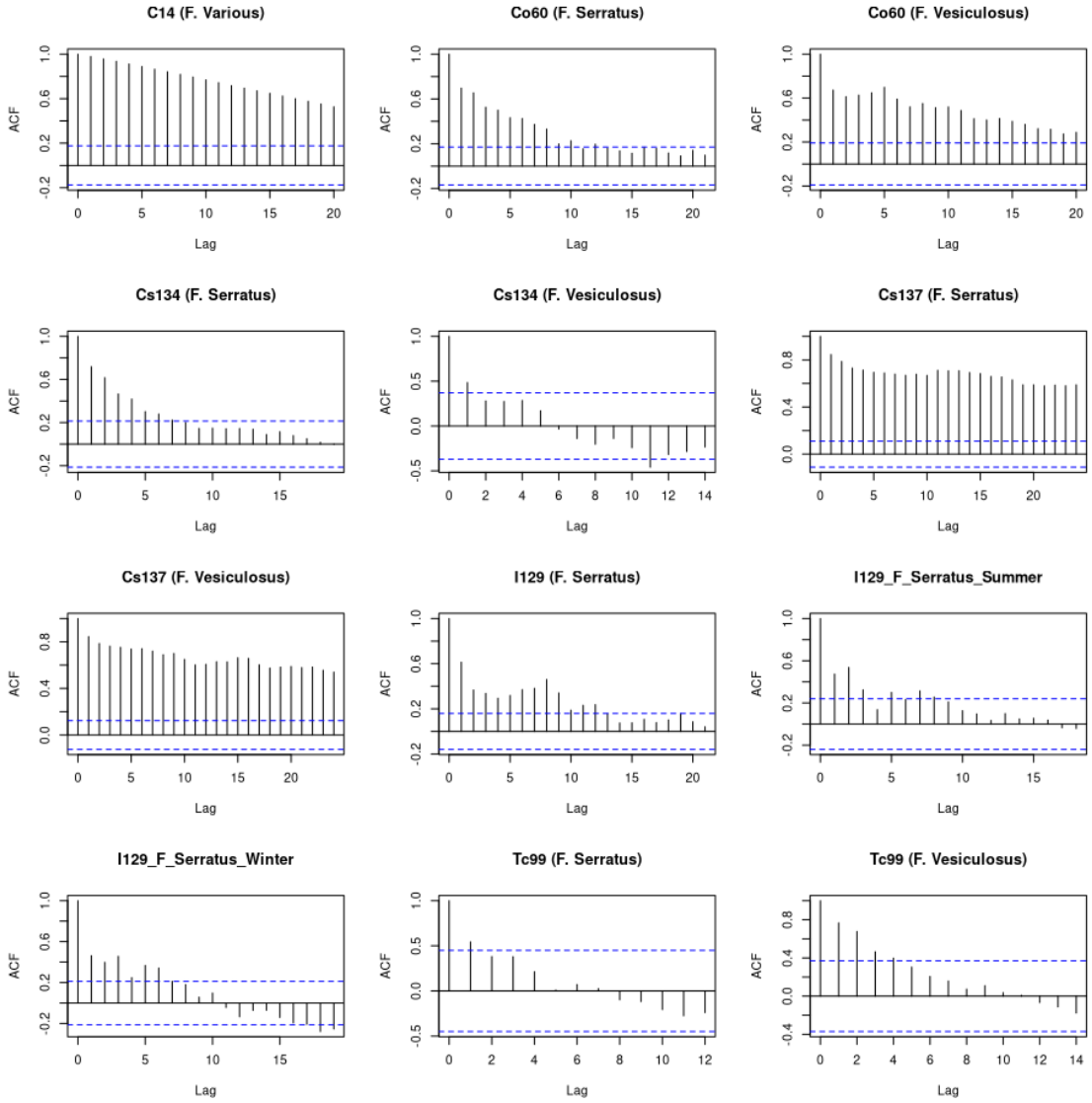


Figure 14: Autocorrelation functions for the measurement data.

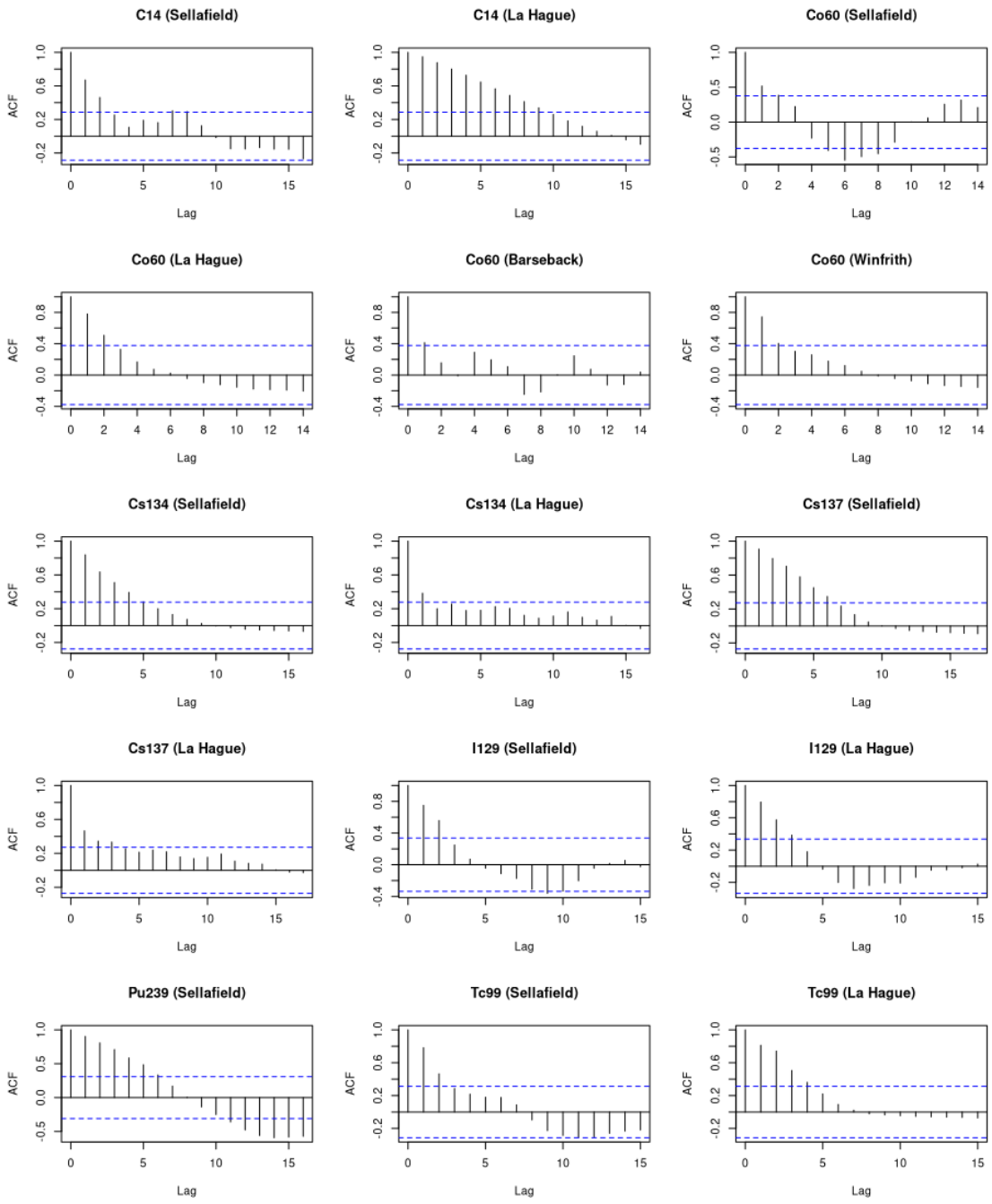


Figure 15: Autocorrelation functions for the marine discharge sources.

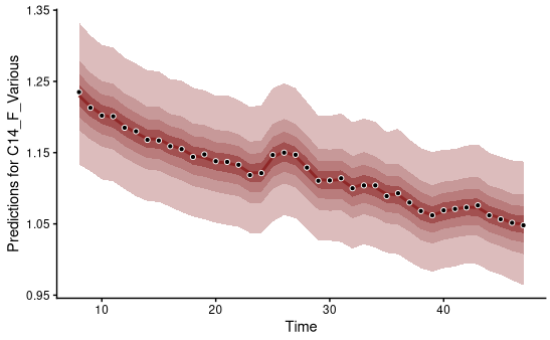
B Baltic Sea outflow simulation

The values here are based on the study performed by Herrmann et al. [11] (pp. 21-22). White noise is added to the process.

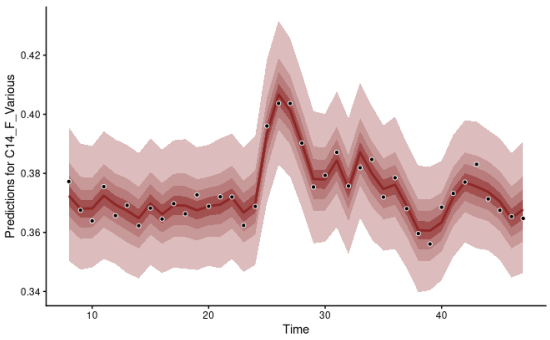
```r

simulate_baltic_cs137 <- function(data_Cs137) {

  # take a vector of years from the original dataset
  years <- sort(unique(data_Cs137$year))
  total_years <- length(years)
  # number of data points after the incident
  n_after <- length(years[years >= 1986])

  t <- seq_len(n_after) - 1

  # effective half lives used from HELCOM 2009 pp. 21-22
  k_1 <- log(2)/2.5 # 1986-1988
  k_2 <- log(2)/9 # 1990 - end
  y_0 <- 650 # approximate peak value Bq / m^-3

  # prefill empty array
  x <- array(5, total_years)

  # expoenential decay with piecewise half lives
  data <- y_0 * exp(- (k_1 * pmin(t, 3) + k_2 * pmax(0, t - 3)))

  # replace the tail (from 1986 to end) by sim data
  x[(total_years - n_after + 1):total_years] <- data

  # add noise
  set.seed(123)
  x <- x + rnorm(total_years, sd = 0.1 * mean(x))
  return(as.vector(x))
}
```

Listing 1: Function to simulate Baltic Cs-137 outflow.

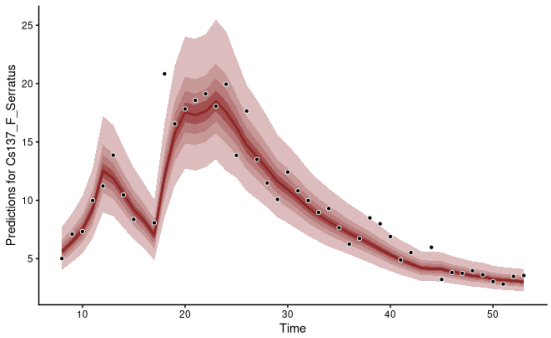
C Full fitted time series



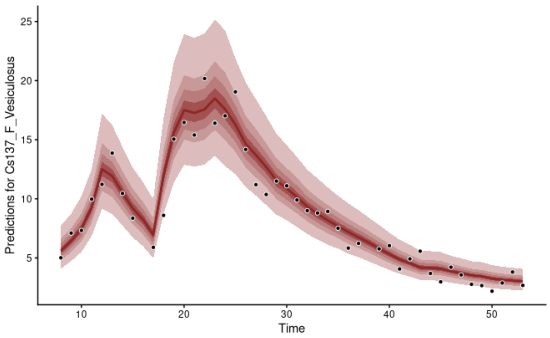
(a) Carbon-14 (AR-based model)



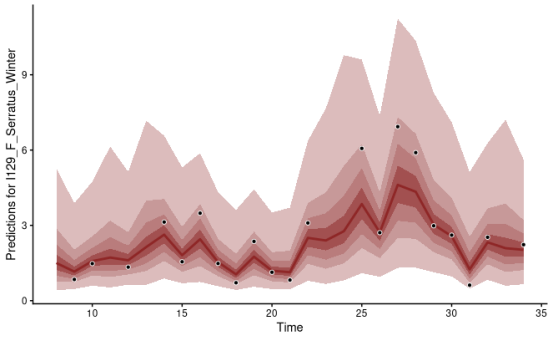
(b) Carbon-14 (Bomb pulse corrected model)



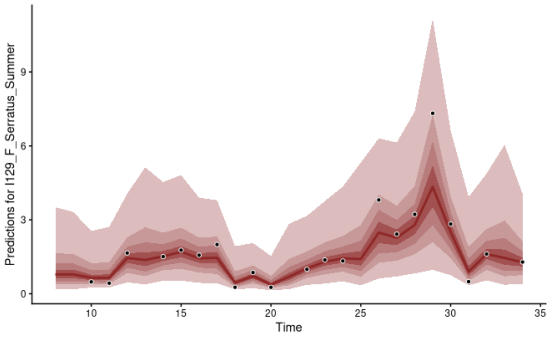
(c) Caesium-137 (F. Serratus)



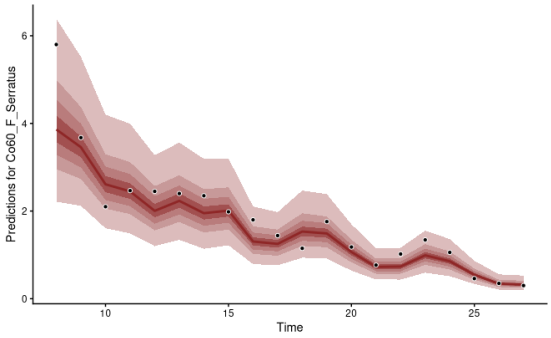
(d) Caesium-137 (F. Vesiculosus)



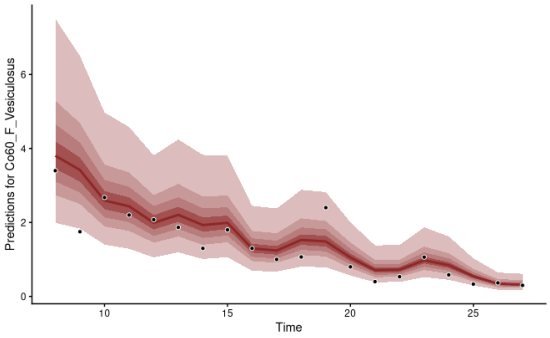
(e) Iodine-129 (Winter)



(f) Iodine-129 (Summer)



(g) Cobalt-60 (F. Serratus)



(h) Cobalt-60 (F. Vesiculosus)