F-boxes for filtering

Olivier Strauss[1] Sebastien Destercke[1,2]

[1]LIRMM (CNRS & Univ. Montpellier II), 161 rue Ada, F-34392 Montpellier cedex 5, France
[2]INRA/CIRAD, UMR1208, 2 place P. Viala, F-34060 Montpellier cedex 1, France

Abstract

Selecting a particular summative kernel (i.e., formally equivalent to a probability distribution) when filtering a digital signal can be a difficult task. An obvious solution to this problem is to filter with multiple kernels rather than with a single one, but the computing cost of such a solution can quickly become prohibitive (especially in real-time applications). Another alternative, the one studied in this paper, is to consider kernels modeled by imprecise probabilistic representations. Considering such representations makes the use of numerical tools coming from imprecise probability theory possible, such as the Choquet integral, and allows one to work with multiple kernels without multiplying too much the number of required computations. In this paper, we propose to use the well-known p-box representation to filter a digital signal. We show that the use p-boxes allows making more precise inferences than those obtained with possibility distributions and clouds. We then discuss the practical aspect of computing a filtered signal with p-boxes, and finish by some experiments.

Keywords: Possibility measures, maxitive kernels, p-boxes, clouds, signal filtering.

1. Introduction

Reconstructing a continuous signal from a set of sampled and possibly corrupted observations is a common problem in both digital analysis and signal processing [1]. In this context, kernel-based methods can be used for different purposes: reconstruction, impulse response modelling, interpolation, linear and non-linear transformations, stochastic or band-pass filtering, etc.

Most of the kernels used in signal processing are summative kernels, or linear combination of summative kernels. A summative kernel is a positive function with an integral equal to one. A summative kernel can therefore be associated to a particular probability distribution. However, how to choose a particular kernel and its parameters to filter a given signal is often a tricky question. In this paper, we focus more particularly on symmetric and unimodal kernels, as they are often used in practice.

One solution to overcome this difficulty is to use imprecise probabilistic models to model in a single representation a whole set of possible kernels. This can be done thanks to the correspondence between sets of summative kernels and sets of probabilities. Existing propositions [2] have explored the use of representations such as possibility distributions [3, 4] or clouds [5] to derive respectively so-called maxitive and cloudy kernels.

These representations have the important advantage of being computationally tractable and are particularly interesting when the kernels to model all share a common modal value.

This paper is focused on another kernel based on an imprecise probabilistic representation, namely p-boxes. Indeed, the interest of using p-boxes for filtering purposes have not been investigated yet, despite the fact that they present the same advantages as clouds and possibility distributions (simplicity and tractability). P-boxes consist in pairs of cumulative distributions, and have the same computational advantages as clouds or possibility distributions. We will show in this paper that they can also provide more precise results than maxitive or cloudy kernels.

Filtering with such representations require to use imprecise extensions of expectation operators, which result in an an interval-valued filtered signal rather than a classical single-valued one. This signal corresponds to the lower and upper envelopes of filtered signals that would have been obtained by filtering with every single summative kernel contained in the imprecise probabilistic representation. Using such representations is therefore particularly interesting to achieve a robustness analysis in one operation.

Section 2 presents the material concerning summative kernels and imprecise probabilistic representations needed in this paper. The capacity of p-boxes to model symmetric kernels and their relation with other representations is studied in Section 3. Section 4 deals with the computational aspects of p-box based filtering. Finally, Section 5 presents some basic experiments on simulated noisy signals.

2. Summative and imprecise kernels

In this section, we recall the basics about summative kernels, before introducing kernels based on imprecise probabilistic representations. We restrict ourselves to representations defined on the real line $\mathbb{R}$ and its discretization $\mathscr{X}$.

2.1. Summative kernels and probability sets

A *summative kernel* κ is formally equivalent to a Lebesgue-measurable probability distribution $\kappa : \mathbb{R} \to \mathbb{R}^+$ and can be interpreted as such. The associated probability measure $P_\kappa : \mathscr{B} \to [0, 1]$ defined on the real Borel agebra $\mathscr{B}$ is such that, for any measurable subset $A \subseteq \mathbb{R}$ (also called an event), $P_\kappa(A) = \int_A \kappa(x)dx$. Some typical summative kernels are recalled and represented in Table 1.
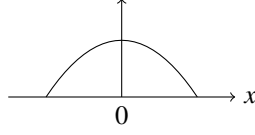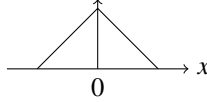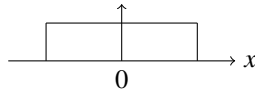
Name	κ	Shape
Epanechnikov	$\kappa(x) = \frac{1}{\Delta}\frac{3}{4}(1-(\frac{x}{\Delta})^2)I_\Delta$	
Triangular	$\kappa_\Delta(x) = (1-|\frac{x}{\Delta}|)I_\Delta$	
Uniform	$\kappa_\Delta(x) = \frac{1}{2\Delta}I_\Delta$	

Table 1: Some classical summative kernels

As said in the introduction, the choice of a particular summative kernel to filter a given signal can be tricky. This is why kernels based on imprecise probability representations can help in filtering with sets of kernels at once.

We call *imprecise* kernel a convex set $\mathscr{P}$ of summative kernels. Since summative kernels are equivalent to probability distributions, $\mathscr{P}$ is equivalent to a convex set of probability distributions. Similarly to the probability measure of a summative kernel, to any event $A \subseteq \mathbb{R}$ can be associated two measures, i.e., the lower and upper probability measures induced by $\mathscr{P}$, which are such that

$$\underline{P}(A) = \inf_{\kappa \in \mathscr{P}} P(A) \text{ and } \overline{P}(A) = \sup_{\kappa \in \mathscr{P}} P(A).$$

These two measures are dual, in the sense that $\underline{P}(A) = 1 - \overline{P}(A^c)$, with A^c the complement of A. Hence it is sufficient in practice to work with only one of them specified over all events. Handling a generic imprecise kernel $\mathscr{P}$ can be a difficult task, however simpler imprecise probabilistic representations that satisfy some properties can make this handling much easier. We introduce such easy-to-use representations in the next sections.

2.2. Possibility distributions and maxitive kernels

A *maxitive kernel* π is defined as a possibility distribution [3], that is a mapping $\pi : \mathbb{R} \to [0,1]$ with at least one $x \in \mathbb{R}$ such that $\pi(x) = 1$. It induces two (lower and upper) confidence measures, respectively called necessity and possibility measures. They are such that, for any event $A \subseteq \mathbb{R}$, we have:

$$\Pi(A) = \max_{x \in A} \pi(x) \qquad N(A) = 1 - \Pi(A^c). \quad (1)$$

A maxitive kernel π induces an imprecise kernel $\mathscr{P}_\pi$ including all summative kernels dominated by the possibility measure Π, i.e., $\mathscr{P}_\pi = \{\kappa \in \mathbb{P}_\mathbb{R} | \forall A \subseteq \mathbb{R}, P(A) \leq \Pi(A)\}$, with $\mathbb{P}_\mathbb{R}$ the set of all summative kernels over $\mathbb{R}$. If a summative kernel κ is in $\mathscr{P}_\pi$, we say by a small abuse of language that π includes κ. This interpretation makes maxitive kernels instrumental tools to filter signals when the identification of a single summative kernel is difficult.

2.3. Clouds and cloudy kernels

Maxitive kernels are very simple imprecise kernels, since they consist in a single mapping over elements $x \in \mathbb{R}$. While this makes them very easy to build and to compute with, it also limits their expressiveness. For instance, they cannot model sets of summative kernels whose bandwidth is both lower- and upper-bounded.

Clouds [5] are a slightly more complex representation with increased expressiveness, but they remain simple enough so that computing with them remains tractable. A *cloudy kernel* or cloud is defined as a pair of mappings $[\pi, \eta]$ from $\mathbb{R}$ to the unit interval $[0,1]$ such that $\eta \leq \pi$ and there is at least one element $x \in \mathbb{R}$ such that $\pi(x) = 1$ and one element $y \in \mathbb{R}$ such that $\eta(y) = 0$.

A cloudy kernel $[\pi, \eta]$ induces an imprecise kernel $\mathscr{P}_{[\pi,\eta]}$ such that

$$\mathscr{P}_{[\pi,\eta]} = \{\kappa \in \mathbb{P}_\mathbb{R} | P_\kappa(\eta_\alpha) \leq 1 - \alpha \leq P_\kappa(\pi_{\overline{\alpha}})\}, \quad (2)$$

where $\eta_\alpha = \{x | \eta(x) \geq \alpha\}$ and $\pi_{\overline{\alpha}} = \{x | \pi(x) > \alpha\}$. Maxitive kernels are special kinds of cloudy kernels, as a maxitive kernel π is equivalent to a cloudy kernel $[\pi, \eta]$ where $\eta = 0$. Also note that, mathematically speaking, clouds are equivalent to interval-valued fuzzy sets having boundary conditions (i.e., $\pi(x) = 1$ and $\eta(y) = 0$ for some $(x,y) \in \mathbb{R}^2$). A family of clouds that will be of particular interest here are the comonotonic clouds [6]. They are defined as follows:

Definition 1. A cloud is comonotonic if $\forall x, y \in \mathbb{R}$, $\pi(x) < \pi(y) \Rightarrow \eta(x) \leq \eta(y)$

Indeed, such clouds possess the properties [6] that makes cloudy kernels tractable representations.

2.4. P-boxes and f-boxes

Let us now introduce the imprecise kernels based on p-boxes, that we call *f-boxes* (for filter boxes).

A f-box $[\underline{F}, \overline{F}]$ is formally equivalent to a p-box [7, 8], and is defined as a pair of cumulative distributions, a lower one $\underline{F} : \mathbb{R} \to [0,1]$ and an upper one $\overline{F} : \mathbb{R} \to [0,1]$, such that $\underline{F} \leq \overline{F}$ and there is at least one element $x \in \mathbb{R}$

with $\underline{F}(x) = \overline{F}(x) = 1$. A f-box induces an imprecise kernel $\mathscr{P}_{[\underline{F},\overline{F}]}$ defined as follows

$$\mathscr{P}_{[\underline{F},\overline{F}]} = \{\kappa \in \mathbb{P}_{\mathbb{R}} | \forall x \in \mathbb{R},\ \underline{F}(x) \le P_{\kappa}([-\infty,x]) \le \overline{F}(x)\}.$$

A f-box encompasses every summative kernel whose cumulative distribution lies between $\underline{F}$ and $\overline{F}$.

3. Modeling symmetric kernels with imprecise kernels

As said in the introduction, in this paper we are mainly concerned by imprecise kernels modeling sets of kernels that are symmetric and unimodal. We assume that the shape of the kernels to model is known and parametrised by a bandwidth (i.e., the support of the corresponding probability). We also assume that this bandwidth is ill-known, and that it is only lower and upper bounded.

We therefore assume from now one that κ belongs to a family of symmetric kernel centered around zero (i.e., $\kappa(x) = \kappa(-x)$). A basic symmetric kernel κ has $[-1,1]$ for support. From this basic kernel it is easy to derive a summative kernel κ_Δ with bandwidth Δ and support $[-\Delta, \Delta]$ using the transformation $\kappa_\Delta(x) = \frac{1}{\Delta}\kappa(\frac{x}{\Delta})$ (see Table 1 for illustrations).

3.1. Maxitive and cloudy kernels modeling: a reminder

First, recall that the cumulative distribution $F_{\kappa_\Delta} : [-\Delta, \Delta] \to [0,1]$ of κ_Δ is such that $F_{\kappa_\Delta}(0) = 1/2$ and $F(x) + F(-x) = 1$.

Let us now briefly remind how a set of kernels κ_Δ whose bandwidth is only known to lie in the interval $[\Delta] := [\Delta_{\inf}, \Delta_{\sup}]$ can be modeled by maxitive or cloudy kernels. We will denote by $\kappa_{[\Delta]}$ this set of kernels.

There are many ways to build a maxitive kernel including the set $\kappa_{[\Delta]}$ [9], but the most specific solution is given by the Dubois-Prade transformation [10]. For $\kappa_{[\Delta]}$, this transformation gives the distribution $\pi_{[\Delta]}$ such that

$$\pi_{[\Delta]}(x) = \begin{cases} 2 * F_{\Delta_{\sup}}(x) & \text{if } x \le 0 \\ 2 * (1 - F_{\Delta_{\sup}}(x)) & \text{if } x > 0. \end{cases} \quad (3)$$

We will denote by $\pi^+_{[\Delta]}, \pi^-_{[\Delta]}$ the following functions

$$\pi^-_{[\Delta]}(x) = \begin{cases} \pi_{[\Delta]}(x) & \text{if } x \le 0 \\ 1 & \text{if } x > 0 \end{cases} \quad \pi^+_{[\Delta]}(x) = \begin{cases} 1 & \text{if } x \le 0 \\ \pi_{[\Delta]}(x) & \text{if } x > 0. \end{cases} \quad (4)$$

The induced imprecise kernel $\mathscr{P}_{\pi_{[\Delta]}}$ includes, among others, all summative kernels $\kappa_{\Delta'}$ with $\Delta' \in [0, \Delta_{\sup}]$ [9]. This means that maxitive kernels alone cannot take account of the lower bound $\Delta_{\inf}$, and include for example the Dirac function. This can be a problem, for instance in those cases where it is desirable to smoothen a signal or where the interval-valued signal resulting from an imprecise filtering should not envelope the initial noisy signal.

By supplementing the (upper) distribution π with a (lower) distribution η, cloudy kernels can solve this insufficiency. In [11], we proposed the following cloudy
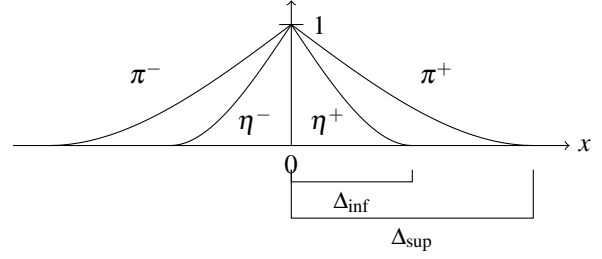

Figure 1: Example of cloudy kernel

$---\ \overline{F} = \max\{F_{\Delta_{\sup}}, F_{\Delta_{\inf}}\}$
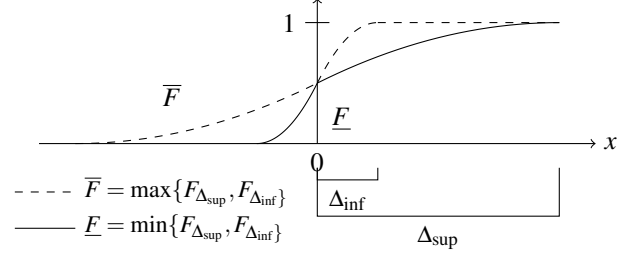$———\ \underline{F} = \min\{F_{\Delta_{\sup}}, F_{\Delta_{\inf}}\}$

Figure 2: Example of f-box

kernel $[\pi_{[\Delta]}, \eta_{[\Delta]}]$ to approximate $\kappa_{[\Delta]}$:

$$\pi_{[\Delta]}(x) = \begin{cases} 2 * F_{\Delta_{\sup}}(x) & \text{if } x \le 0 \\ 2 * (1 - F_{\Delta_{\sup}}(x)) & \text{if } x \ge 0 \end{cases} \quad (5)$$

$$\eta_{[\Delta]}(x) = \begin{cases} 2 * F_{\Delta_{\inf}}(x) & \text{if } x \le 0 \\ 2 * (1 - F_{\Delta_{\inf}}(x)) & \text{if } x \ge 0 \end{cases} \quad (6)$$

Similarly to maxitive kernels, we denote by $\eta^+_{[\Delta]}, \eta^-_{[\Delta]}$ the functions such that

$$\eta^-_{[\Delta]}(x) = \begin{cases} \eta_{[\Delta]}(x) & \text{if } x \le 0 \\ 1 & \text{if } x > 0 \end{cases} \quad \eta^+_{[\Delta]}(x) = \begin{cases} 1 & \text{if } x \le 0 \\ \eta_{[\Delta]}(x) & \text{if } x > 0. \end{cases} \quad (7)$$

A cloudy kernel corresponding to a family of triangular kernel κ is pictured in Figure 1 (the corresponding maxitive kernel is obtained by suppressing η). It can be shown that the cloudy kernel $[\pi_{[\Delta]}, \eta_{[\Delta]}]$ includes kernels κ_Δ with $\Delta \in [\Delta]$ (including $\kappa_{\Delta_{\inf}}$ and $\kappa_{\Delta_{\sup}}$), and no kernels κ_Δ such that $\Delta \notin [\Delta]$. Cloudy kernels therefore allow to solve some issues related to maxitive kernels (but have problems of their own, as we shall see in Section 5).

3.2. F-box kernels

As illustrated by Figure 1, both maxitive and cloudy kernels are symmetric and are built around the modal value (here, 0). Rather than considering this modal value as a central point, the f-box approach focuses on events of the kind $[-\infty, x]$. Such a view suggests different approximations of the set $\kappa_{[\Delta]}$, that we explore here.

Given the set of kernels $\kappa_{[\Delta]}$, the most specific f-box $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ one can build that encompasses this set is

$$\overline{F}_{[\Delta]}(x) = \max\{F_{\Delta_{\sup}}(x), F_{\Delta_{\inf}}(x)\}, \quad (8)$$

$$\underline{F}_{[\Delta]}(x) = \min\{F_{\Delta_{\sup}}(x), F_{\Delta_{\inf}}(x)\}, \quad (9)$$

as shows the next theorem:

937

Theorem 1. $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ *is the most specific f-box including the set of kernels* $\kappa_{[\Delta]}$, *and does not include any kernel* κ_Δ *with* $\Delta \notin [\Delta]$.

Proof. (**Sketch**) Consider a kernel κ_Δ with $\Delta \in [\Delta]$. It is included in $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$, as

$$\min\{F_{\Delta_{\sup}}, F_{\Delta_{\inf}}\} \le F_\Delta \le \max\{F_{\Delta_{\sup}}, F_{\Delta_{\inf}}\},$$

meaning that both $\kappa_{\Delta_{\sup}}$ and $\kappa_{\Delta_{\inf}}$ are in the kernel. To show that it is the most specific, assume *ex absurdo* that there is a more specific f-box $(\underline{F}', \overline{F}')$ including $\kappa_{[\Delta]}$. This means that $\exists x$ such that $\underline{F}'(x) > \underline{F}_{[\Delta]}(x)$ or $\overline{F}'(x) < \overline{F}_{[\Delta]}(x)$, implying that either $F_{\Delta_{\sup}} \notin (\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ or $F_{\Delta_{\inf}} \notin (\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$, which is a contradiction since both $\kappa_{\Delta_{\sup}}$ and $\kappa_{\Delta_{\inf}}$ are in $\kappa_{[\Delta]}$. To show that no kernel κ_Δ with $\Delta \notin [\Delta]$ is in $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$, simply takes the same argument (there is an x for which $F_\Delta(x) < \underline{F}_{[\Delta]}(x)$ or $F_\Delta(x) > \overline{F}_{[\Delta]}(x)$). $\square$

Figure 2 shows a f-box given by Eq. (8) and (9) for triangular kernels. We can see in this picture that, although there is some symmetry in the representation, it is not symmetric around the 0 coordinate axis.

The next theorem shows that f-boxes modeling symmetric unimodal kernels are more specific than maxitive or cloudy kernels given by Eq. (3) and (6), and will therefore provide more precise results when filtering.

Theorem 2. $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ *is strictly more specific than the maxitive kernel* $\pi_{[\Delta]}$ *or the cloudy kernel* $[\pi_{[\Delta]}, \eta_{[\Delta]}]$, *in the sense that* $\mathscr{P}_{(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})} \subseteq \mathscr{P}_{[\pi_{[\Delta]}, \eta_{[\Delta]}]} \subseteq \mathscr{P}_{\pi_{[\Delta]}}$, *with the inclusion being usually strict.*

Proof. (**Sketch**) In the case of symmetric unimodal kernels around 0, imprecise kernels induced by maxitive kernel $\pi_{[\Delta]}$ and cloudy kernel $[\pi_{[\Delta]}, \eta_{[\Delta]}]$ correspond to sets of summative kernels that satisfy lower and upper probabilistic contraints induced by $\pi_{[\Delta]}$ and $[\pi_{[\Delta]}, \eta_{[\Delta]}]$ around events $[-x, x]$ (see [5] and [12] for details).

To show that $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ is more precise than both of them, it is then sufficient to show that lower and upper probabilities induced by $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ over events $[-x, x]$ are tighter or coincide with the one induced by $\pi_{[\Delta]}$ and $[\pi_{[\Delta]}, \eta_{[\Delta]}]$.

Using Eq. (5) and (6), we get that $\underline{P}_{\pi_{[\Delta]}}([-x, x]) = \underline{P}_{[\pi_{[\Delta]}, \eta_{[\Delta]}]}([-x, x]) = 1 - 2F_{\Delta_{\sup}(-x)} = F_{\Delta_{\sup}}(x) - F_{\Delta_{\sup}}(-x)$ and $\overline{P}_{[\pi_{[\Delta]}, \eta_{[\Delta]}]}([-x, x]) = 1 - 2F_{\Delta_{\inf}(-x)} = F_{\Delta_{\inf}}(x) - F_{\Delta_{\inf}(-x)}$ (the upper probability bound for $\pi_{[\Delta]}$ on such events is one).

Figure 3 shows that the lower probability $\underline{P}_{(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})}$ has $F_{\Delta_{\sup}(x)} - F_{\Delta_{\sup}(-x)}$ for upper bound (in fact, this is the exact value). The same reasoning can be done for the upper probability.

To see that the inclusion is usually strict, remark that the kernel which has function $\pi_{[\Delta]}^-$ for cumulative distribution is not in $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$, while it is in $\pi_{[\Delta]}$ and $[\pi_{[\Delta]}, \eta_{[\Delta]}]$. $\square$

Finally, we can show that any kernel whose cumulative distributions lies in $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ has a bandwidth in $[\Delta]$
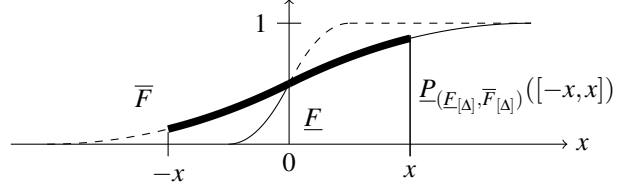

Figure 3: Lower probability on event $[-x, x]$

Theorem 3. *Any summative kernel* $\kappa \in \mathscr{P}_{(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})}$ *has a bandwidth in* $[\Delta]$.

Proof. (**Sketch**) First consider the minimal bandwidth $\Delta_{\inf}$. Since for all $x \in [-\Delta_{\inf}, \Delta_{\inf}]$, $0 < \underline{F}_{[\Delta]}(x) < \overline{F}_{[\Delta]}(x) < 1$, no cumulative distribution in $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ can have a bandwidth lower than $\Delta_{\inf}$.

Consider now $\Delta_{\sup}$. As $\overline{F}_{[\Delta]}(x) = 0$ for any $x < -\Delta_{\sup}$ and $\overline{F}_{[\Delta]}(x) = 1$ for any $x > \Delta_{\sup}$, no cumulative distribution in $(\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]})$ can have a bandwidth higher than $\Delta_{\sup}$. $\square$

4. Practical computations

We can now discuss how filtering can actually be achieved with f-boxes while maintaining a low computational complexity. This can be done by extending classical expectation operators (equivalent to filter with summative kernels) through the use of Choquet integrals. We first recall what is a Choquet integral [13, 14] and its links with expectation operators, summative kernels and maxitive kernels. We then propose an efficient algorithm allowing to compute this Choquet integral for f-boxes. To make notations shorter, we will from now on denote $\underline{F}_{[\Delta]}, \overline{F}_{[\Delta]}$ by $\overline{F}$ and $\underline{F}$, assuming that a kernel family as well as maximal and minimal bandwidth have been specified.

Since computations are achieved on a discretised space, we consider in this section that we are working on a finite domain $\mathscr{X}$ of N elements. In our case, this space corresponds to a finite sampling of the (continuous) signal.

4.1. Expectation operator and Choquet integral

Let $\mathscr{X} = \{x_1, \ldots, x_N\}$ be the discretized domain, where x_1 corresponds to the last sampled time and x_N is the sampled item with time horizon N. Consider a real-valued function f (here, the sampled values of the signal) on $\mathscr{X}$, together with a discretized summative kernel $\kappa_i, i = 1, \ldots, N$, where $\kappa_i = \kappa(x_i)$.

Classical causal convolution between the kernel κ and the sampled signal f is equivalent to apply an expectation operator, i.e. to compute $\mathbb{E}_\kappa(f)$ such that

$$\mathbb{E}_\kappa(f) = \sum_{i=1}^{N} \kappa_i f(x_i)$$

Now, when working with a set $\mathscr{P}$ of kernels defined on $\mathscr{X}$, expectation operator $\mathbb{E}$ becomes imprecise, and

its result when applied to f is an interval-valued expectation $[\underline{\mathbb{E}}(f), \overline{\mathbb{E}}(f)]$ such that

$$\underline{\mathbb{E}}(f) = \inf_{\kappa \in \mathscr{P}} \mathbb{E}_\kappa(f) \qquad \overline{\mathbb{E}}(f) = \sup_{\kappa \in \mathscr{P}} \mathbb{E}_\kappa(f) \qquad (10)$$

In general, these bounds are not easy to compute. However in some specific cases practical tools exists that makes them computable. In particular, if there is no loss of information when considering the set of summative kernels dominated by the lower probability of $\mathscr{P}$ on events alone, and if the lower probability $\underline{P}$ satisfies a property of 2-monotinicity, that is if for any pair $\{A, B\} \subseteq \mathscr{X}$ we have $\underline{P}(A \cap B) + \underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$, then Eq. (10) can be solved by using the Choquet integral.

Consider a positive bounded function[1] f on $\mathscr{X}$. If we denote by $()$ a reordering of elements of $\mathscr{X}$ such that $f(x_{(1)}) \leq \ldots \leq f(x_{(N)})$, Choquet Integrals giving lower and upper expectations are given by

$$C_{\underline{P}}(f) = \underline{\mathbb{E}}(f) = \sum_{i=1}^{N} (f(x_{(i)}) - f(x_{(i-1)})) \underline{P}(A_{(i)}),$$

$$C_{\overline{P}}(f) = \overline{\mathbb{E}}(f) = \sum_{i=1}^{N} (f(x_{(i)}) - f(x_{(i-1)})) \overline{P}(A_{(i)}),$$

with $f(x_{(0)}) = 0$ and $A_{(i)} = \{x_{(i)}, \ldots, x_{(N)}\}$. The main difficulty then becomes to compute the lower and upper confidence measures for the N sets A_i.

4.2. Imprecise expectations with f-boxes

Since p-boxes, or equivalently f-boxes induce lower confidence measures that are ∞-monotone [8] (an even more restrictive property than 2-monotonicity), Choquet integral can be used to compute their expectation bounds. Let us now see how the lower confidence measure value on various events can be computed efficiently (upper confidence measure values can then be computed by duality).

In the continuous case, when a f-box is defined by two continuous fucntions $\overline{F}$ and $\underline{F}$ over $\mathbb{R}$, the lower probability $\underline{P}([x, y])$ of an interval $C = [x, y]$ is given by the following formula

$$\underline{P}(C) = \max\{0, \underline{F}(y) - \overline{F}(x)\},$$

and the lower probability of the union A of any set of n disjoint intervals $C_i = [x_i, y_i]$, $i = 1, \ldots, n$ is given by

$$\underline{P}\left(A = \bigcup_{i=1}^{n} C_i\right) = \sum_{i=1}^{n} \underline{P}(C_i). \qquad (11)$$

Note that any event A can be written in this form.

In the discretised case where $\mathscr{X} = \{x_1, \ldots, x_N\}$, an interval $C = [x, y]$ becomes a set $C = [x_i, x_j]$ where $C = \{x_k | i \leq k \leq j\}$ contains every discretised elements x_k

[1]Assuming positivity is not constraining here, since if c is a constant $\underline{\mathbb{E}}(f + c) = \underline{\mathbb{E}}(f) + c$ and the same holds for $\overline{\mathbb{E}}$. Therefore any bounded function can be made positive by a simple translation.

between x_i and x_j. In this case, the lower probability of C becomes

$$\underline{P}(C) = \max\{0, \underline{F}(x_j) - \overline{F}(x_{i-1})\},$$

and Eq. (11) remains the same, and again any event A can be written as the union of disjoint intervals (two intervals $[x_i, x_j], [x_k, x_l]$ are considered disjoint if $j + 1 < k$, otherwise if $k = j + 1$ their union is no longer disjoint).

Algorithm 1 describes how to compute lower confidence measures and the incremental summation giving the lower expectation, while Example 1 provides an illustration of the process. At each step, the disjoint intervals of the form $C_1, \ldots, C_{M_i}$ forming $A_{(i)}$ are extracted and the corresponding lower confidence measure is computed. The value of the Choquet integral is then incremented. Note that two orderings and indexing are used in the algorithm: the one where elements are ordered by values of f, denoted by $()$, and the other where elements are ordered w.r.t. to their sampling time ($f(x_1)$ being the first sampled value and $f(x_N)$ the last one).
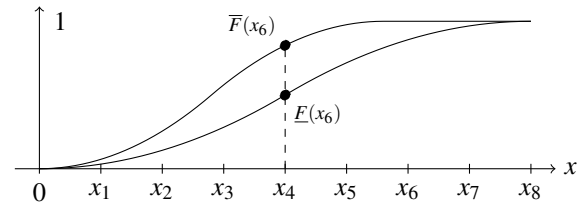

Figure 4: Discretisation of a f-filter with 8 sampled values

Algorithm 1: Algorithm for lower expectations: basic ideas

Input: $f, [\underline{F}, \overline{F}]$, N (number of discretized points)
Output: Lower/upper expectations
$\underline{\mathbb{E}} = 0$;
for $i = 1, \ldots, N$ **do**
  Compute $f(x_{(i)}) - f(x_{(i-1)})$;
  Extract intervals $C_1, \ldots, C_{M_i}$ such that
  $A_{(i)} = C_1 \cup \ldots \cup C_{M_i}$;
  With $C_j = [x_{\underline{j}}, x_{\overline{j}}]$;
  Compute
  $\underline{P}(A_{(i)}) = \sum_{j=1}^{M_i} \max(0, \underline{F}(x_{\overline{j}}) - \overline{F}(x_{\underline{j}-1}))$;
  $\underline{\mathbb{E}} = \underline{\mathbb{E}} + (f(x_{(i)}) - f(x_{(i-1)}) \times \underline{P}(A_{(i)})$

Example 1. Consider the situation pictured in Fig. 4. For each sampled item x and associated sampled value $f(x)$, we consider that the corresponding distribution values are $[\underline{F}(x), \overline{F}(x)]$ (More conservative discretizations can be adopted, but do not change the problem). The values of discretized $f, \underline{F}$ and $\overline{F}$ are summarized in Table 2.

If we now apply Algorithm 1 with the lower probability, we get the following steps:

1. i=1, $f(x_{(i)}) - f(x_{(i-1)}) = 2$, $A_{(i)} = \mathscr{X}$, $\underline{P}(A_{(i)}) = 1$ $\rightarrow \underline{\mathbb{E}} = 2$

2. i=2, $f(x_{(i)}) - f(x_{(i-1)}) = 3$, $A_{(i)} = C_1 = \{x_2, \ldots, x_8\}$, $\underline{P}(A_{(i)}) = \max\{0, \underline{F}(x_8) - \overline{F}(x_1)\} = 0.936 \rightarrow \underline{\mathbb{E}} = 4.808$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\overline{F}$	0.064	0.255	0.569	0.837	0.977	1	1	1
$\underline{F}$	0.031	0.125	0.281	0.5	0.719	0.875	0.969	1
f	2	9	8	10	5	6	12	13
()	$x_{(1)}$	$x_{(5)}$	$x_{(4)}$	$x_{(6)}$	$x_{(2)}$	$x_{(3)}$	$x_{(7)}$	$x_{(8)}$

Table 2: Values for example 1

3. $i=3$, $f(x_{(i)}) - f(x_{(i-1)}) = 1$, $A_{(i)} = C_1 \cup C_2 = \{x_2,\ldots,x_4\} \cup \{x_6,x_7,x_8\}$, $\underline{P}(A_{(i)}) = \max\{0, \underline{F}(x_4) - \overline{F}(x_1)\} + \max\{0, \underline{F}(x_8) - \overline{F}(x_5)\} = 0.469 \to \mathbb{E} = 5.277$

4. $i=4$, $f(x_{(i)}) - f(x_{(i-1)}) = 2$, $A_{(i)} = C_1 \cup C_2 = \{x_2,\ldots,x_4\} \cup \{x_7,x_8\}$, $\underline{P}(A_{(i)}) = \max\{0, \underline{F}(x_4) - \overline{F}(x_1)\} + \max\{0, \underline{F}(x_8) - \overline{F}(x_6)\} = 0.436 \to \mathbb{E} = 5.277 + 2 \times 0.436 = 6.149$

5. $i=5$, $f(x_{(i)}) - f(x_{(i-1)}) = 1$, $A_{(i)} = C_1 \cup C_2 \cup C_3 = \{x_2\} \cup \{x_4\} \cup \{x_7,x_8\}$, $\underline{P}(A_{(i)}) = \max\{0, \underline{F}(x_2) - \overline{F}(x_1)\} + \max\{0, \underline{F}(x_4) - \overline{F}(x_3)\} + \max\{0, \underline{F}(x_8) - \overline{F}(x_6)\} = 0.061 \to \mathbb{E} = 6.21$

6. $i=6$, $f(x_{(i)}) - f(x_{(i-1)}) = 1$, $A_{(i)} = C_1 \cup C_2 = \{x_4\} \cup \{x_7,x_8\}$, $\underline{P}(A_{(i)}) = \max\{0, \underline{F}(x_4) - \overline{F}(x_3)\} + \max\{0, \underline{F}(x_8) - \pi(x_6)\} = 0 \to \mathbb{E} = 6.21$

7. $i=7$, stop, since $\underline{P}(A_{(i-1)}) = 0$

The lower expectation is finally 6.21.

5. Experiments

This section aims at illustrating the behaviours of the different imprecise kernels presented in Section 2 when filtering a noisy signal. We compare four approaches: the conventional precise method (using only one summative kernel) and the imprecise filtering method performed with maxitive, cloudy and f-box kernels.

5.1. Experiment description

The comparison is achieved on the signal shown in Figure 5. This signal is made of 10 sine signals whose frequencies were randomly chosen (in a manner that subsequent sampling would respect the Shannon sampling theorem [15]), with each signal weighted by a randomly chosen weighting factor. Finally, the signal has been quantized and altered by a Poisson noise (so that the noise is neither ergodic nor time invariant). Figure 5 gives the plot of the noisy signal s (in cyan) and of the original signal f (in black).

Assume that the practitioner has decided that the appropriate kernel shape was of the form

$$\kappa_\Delta(x) = e^{\left(\frac{1}{\frac{x}{\Delta}^2 - 1}\right)},$$

but that he could only "guess" that the appropriate bandwidth Δ would lie somewhere in the interval $[\Delta] = [\Delta_{\inf}, \Delta_{\sup}] = [0.045, 0.050]$. The two extreme kernels
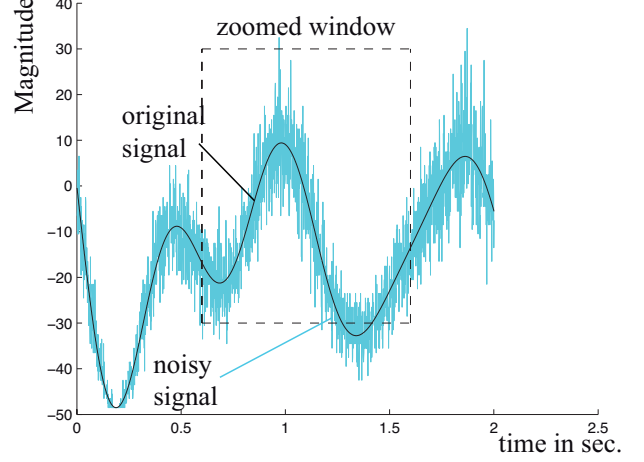

Figure 5: Original signal (black) superimposed on the noisy signal (cyan).
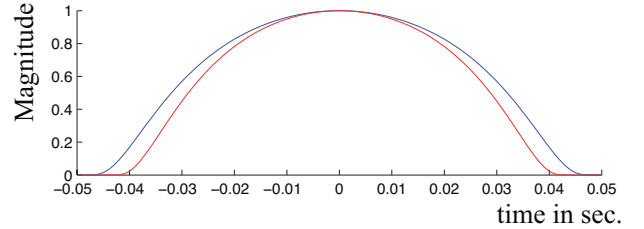

Figure 6: The upper (blue) and lower (red) impulse responses used to define the three capacities.

$\kappa_{0.05}$ (with bandwidth 0.05, in blue) and $\kappa_{0.045}$ (with bandwidth 0.045, in red) are given in Figure 6

The maxitive kernel chosen for this experiment is the most specific that dominates $\kappa_{0.05}$. It is built after Eq. (3) (see [2] for a more generic procedure using the same arguments).

The cloudy kernel is built after Eq. (5) and (6), as explained in [11]. The lower distribution η is built using $\kappa_{0.045}$, while the upper distribution π is built using $\kappa_{0.05}$.

We finally compute a f-box kernel according to the method described in this paper. Using Eq.(8) and (9), we define $\underline{F}$ and $\overline{F}$ as follows. $\forall x \in \mathbb{R}$, let $F_{0.045}(x) = \int_{-\infty}^x \kappa_{0.045}(u)du$ and $F_{0.050}(x) = \int_{-\infty}^x \kappa_{0.050}(u)du$, then $\underline{F}(x) = min(F_{0.045}(x), F_{0.050}(x))$ and $\overline{F}(x) = max(F_{0.045}(x), F_{0.050}(x))$.
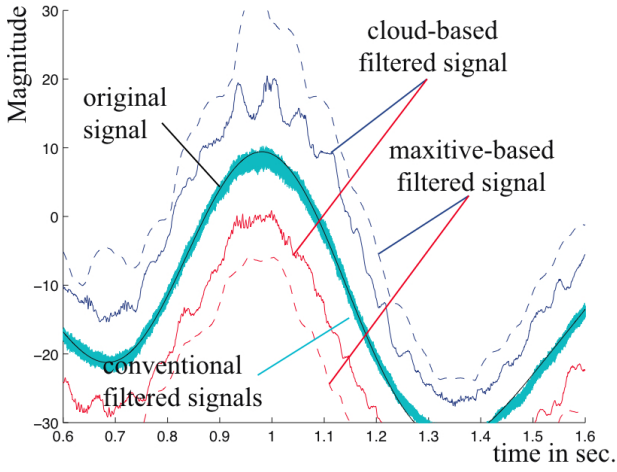
Figure 7: Ten summative-based filtered signals (cyan) superimposed with the maxitive-based filtered signal (dotted blue and red), the cloud-based filtered signal (plain blue and red) and the original signal (black).

5.2. Discussion on experiment results.

The use of imprecise filtering (with maxitive, cloudy or f-box kernels) is recommended in those cases where the practitioner does not know the exact shape of the impulse response of the filter. When the practitioner approximately knows the shape of the filter but can only give its upper bandwidth, then the most appropriate representation is the maxitive kernel. When the practitioner can also give a lower bandwidth, then the cloudy kernel is an appropriate and more specific representation that leads to a more specific interval-valued filtered output. However, the cloudy representation has some theoretical and practical limitations. First, like the maxitive representation, it is built around a modal value and its extension to other situations often provide unsatisfying results. Second, experimental practice shows that it can lead to very chaotic bounds, even if a smoothing filter has been chosen initially by the practitioner. As we shall see here, the f-box representation seems to overcome these problems and leads to a more specific and less chaotic imprecise valued filtered signal. This can be explained by the better properties, in terms of approximation of the set $\kappa_{[\Delta]}$, of f-box kernels.

Figures 7, 8 and 9 display a zoomed window of the resulting filtered output of the different filter. Each Figure presents two imprecise valued filtered signals (blue upper and red lower), the original signal (in black) and the output of 50 summative (conventional) based precise filtered signals involving a kernel of the form κ_Δ with $\Delta \in [0.045, 0.050]$. As expected by the theory (i.e., imprecise kernels all include the set $\kappa_{[\Delta]}$), all the 50 precise filtered signals belong to the enveloppe of each imprecise valued filtered signals. Moreover, since the filter has been correctly designed, $\forall x \in \mathbb{R}$ there is always a value $\Delta \in [0.045, 0.050]$ such that $\mathbb{E}_{\kappa_\Delta}^x(s) = f(x)$, i.e., the exact original signal is retrieved for some value of Δ. Thus, the real signal is always included in all the impre-
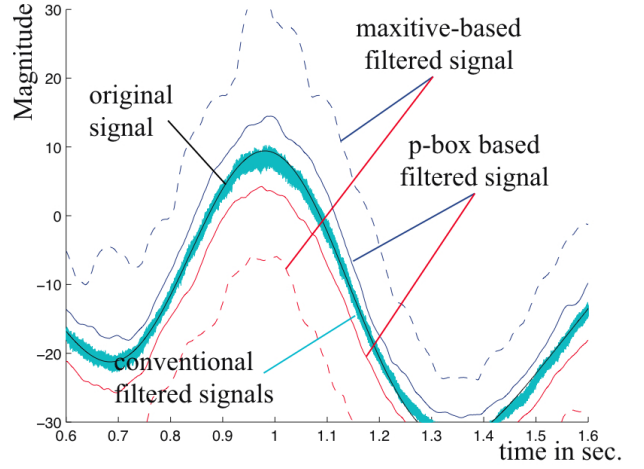

Figure 8: Ten summative-based filtered signals (cyan) superimposed with the maxitive-based filtered signal (dotted blue and red), the cloud-based filtered signal (plain blue and red) and the original signal (black).
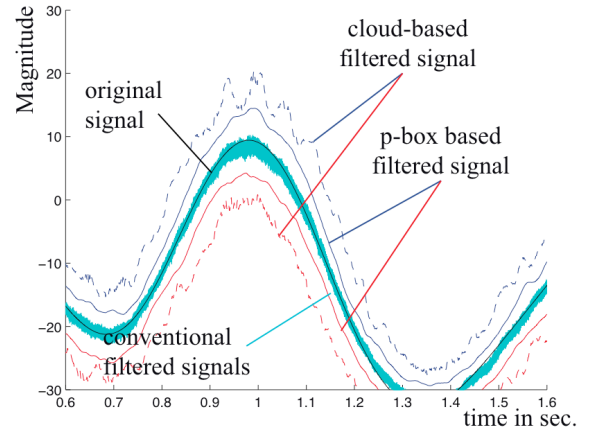

Figure 9: Ten summative-based filtered signals (cyan) superimposed with the maxitive-based filtered signal (dotted blue and red), the cloud-based filtered signal (plain blue and red) and the original signal (black).

cise valued filtered signals.

As expected, the f-box based imprecise valued filtered signal is included in the imprecise valued filtered signals obtained with the maxitive (see Figure 8) and the cloudy kernels (see Figure 9). The inclusion of the cloudy based imprecise valued filtered signal into the maxitive based imprecise valued filtered signal (see Figure 7) has been proved in [11]. Similarly, we proved in this paper that the f-box filtered signal would be included in the two others.

The inclusion of the Dirac measure in the maxitive kernel gives very large upper and lower filtered bounds, that encompass the whole signal (i.e., the signal is always in the interval provided by the maxitive kernel). Compared to maxitive kernels, cloudy kernels are more appropriate to represent the available knowledge on the kernel shape and bandwidth. However, as it can be remarked in Figures 7 and 9, the bounds of the cloudy based imprecise filtered signal are very chaotic and are not similar to the general shape of the precise filtered signals (in cyan on the figures). This provides results that may be difficult to interpret.

On the other hand, the f-box filtered signal is smoother and well follows the general trend of precise signals filtered with different summative kernels. It is also a bit more precise than the cloudy filtered signal. Hence, the use of f-box kernels appears as a promising solution in this kind of situation.

6. Final remarks and perspectives

In this paper, we have introduced f-boxes as convenient models to achieve filtering when the bandwidth of the desired filter is ill-known. F-boxes allow us to model in a single representation a set of (here symmetric and unimodal) kernels whose bandwidth is lower- and upper-bounded. Thanks to their particular properties (i.e., they induce ∞-monotone lower probabilities that possess particular additive properties), we have developed an efficient algorithm, based on the Choquet integral, that computes the lower and upper envelope of sets of filtered signals.

We have also shown that when it comes to model sets of symmetric and unimodal summative kernes, f-boxes display particularly good properties, of theoretical as well as of practical nature. They thus constitute a good approximation and interesting tool to achieve imprecise filtering.

Also, compared to other kernels derived from imprecise probabilistic models, such as maxitive or cloudy kernels, f-boxes are well adapted to the case of causal filtering, mainly because the p-box representation use bounding probabilities over sets $[-\infty, x]$, rather than over sets centred around a modal value. In future works, we intend to explore the potential of f-boxes to model sets of causal filters as well.

We would also like to remark that this model directly extends to the case of positive filters with a fixed gain higher or lower than one. Indeed, since lower and upper expectations are such that $\underline{\mathbb{E}}(a \cdot f) = a \cdot \underline{\mathbb{E}}(f)$ and $\overline{\mathbb{E}}(a \cdot f) = a \cdot \overline{\mathbb{E}}(f)$, it is sufficient to consider the kernel such that $a \cdot \kappa$, where κ is an unknown summative kernel and to do process exactly as before.

References

[1] J. Jan. *Digital Signal Filtering, Analyses and Restoration*. IET, 2000.

[2] K. Loquin and O. Strauss. On the granularity of summative kernels. *Fuzzy Sets and Systems*, 159(1952-1972), 2008.

[3] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.

[4] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.

[5] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.

[6] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: II. Clouds. *International Journal of Approximate Reasoning*, 49(3):664–677, 2008.

[7] S. Ferson, L. Ginzburg, V. Kreinovich, D.M. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.

[8] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: I. Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008.

[9] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.

[10] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In *Proc. of the Fourth International Fuzzy Systems Association World Congress (IFSA'91)*, pages 50–53, Brussels, Belgium, 1991.

[11] Sébastien Destercke and Olivier Strauss. Using cloudy kernels for imprecise linear filtering. In *IPMU*, pages 198–207, 2010.

[12] I. Couso, S. Montes, and P. Gil. The necessity of the strong alpha-cuts of a fuzzy set. *Int. J. on Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:249–262, 2001.

[13] G. Choquet. Theory of capacities. *Annales de l'institut Fourier*, 5:131–295, 1954.

[14] D. Denneberg. *Fuzzy Measures and Integrals – Theory and Applications*, chapter Non-additive measure and integral, basic concepts and their role for applications, pages 42–69. Physica Verlag, 2000.

[15] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.