Received: 22 July 2024 | Revised: 7 September 2024 | Accepted: 12 September 2024

DOI: 10.1111/jfd.14025

RESEARCH ARTICLE

Journal of Fish Diseases WILEY

Classification of *Kudoa thyrsites* infected and uninfected fish using a handheld near-infrared spectrophotometer, SIMCA and PLS-DA

S. Henning

Department of Food Technology and Science, Faculty of Applied Sciences, Cape Peninsula University of Technology, Bellville, South Africa

Correspondence
S. Henning, Department of Food Technology and Science, Faculty of Applied Sciences, Cape Peninsula University of Technology, Bellville, South Africa.
Email: hennings@cput.ac.za

Funding information
National Research Foundation, South Africa

Abstract

Kudoa thyrsites infection of marine fish typically results in myoliquefaction, which is only apparent 24 to 56 h *post-mortem*. The traditional methods for the detection of *K. thyrsites* infected fish are time-consuming and destructive, reducing its marketability. This poses a challenge for the fish industry to remove infected fish before it reaches the market or further processing activities. This study investigated the use of near-infrared (NIR) spectroscopy, in combination with soft independent modelling of class analogy (SIMCA) and partial least square discriminant analysis (PLS-DA), for discriminating *K. thyrsites* infected fish from uninfected fish. Performance of the classification models was evaluated by calculating the sensitivity, specificity and precision. A total of 334 fish samples (200 sardine, 64 hake and 70 kingklip) were used for this study. Infection of *K. thyrsites* was determined with the use of qPCR assays. Ninety per cent (90%) of the sardine samples, 78% of the hake samples and 37% of the kingklip samples were infected. Class groups of infected and uninfected fish samples were created for the purpose of generating SIMCA and PLS-DA classification models for each species of fish, as well as for a species independent data set. Principal component analysis (PCA) of NIR spectra did not show any clustering for infected and uninfected samples. Calibration and test sample sets were generated for the purpose of building and testing the SIMCA and PLD-DA classification models. SIMCA and PLS-DA were unable to classify test samples correctly into the two classes. The number of misclassifications (NMC) was higher for the SIMCA models than for the PLS-DA models, with more than 60% incorrectly classified. SIMCA classified most of the test samples into both classes. The precision for PLS-DA were 89% for sardine, 81% for hake, 0% for kingklip and 87% for species independent models, however, most samples were classified at infected. The use of NIR spectroscopy and classification models such as SIMCA and PLS-DA showed limited use as a method to distinguish between *K. thyrsites* infected and uninfected fish samples. Textural and chemical changes during extended frozen storage of the fish samples may have masked the effects associated with *K. thyrsites* infection. Further studies are suggested where NIR spectroscopy is used in combination with texture analysis and image spectroscopy.

KEYWORDS
Cape hake, kingklip, non-destructive, quality control, sardines

1 | INTRODUCTION

The presence of parasites in fish fillets is a serious quality and food safety concern (Huss, 1994). Although the parasite *Kudoa thyrsites* does not pose a consumer health risk (Alvarez-Pellitero & Sitja-Bobadilla, 1993), the infection of marine fish species by this Myxozoan parasite is associated with quality defects such as the development of *post-mortem* myoliquefaction and unsightly cysts (Dawson-Coates et al., 2003; Gilchrist, 1924; Levsen et al., 2008; Samaranayaka et al., 2006; Webb, 1990). The development of myoliquefaction is a result of a cysteine cathepsin L protease released by *K. thyrsites* (Funk et al., 2008; Tsuyuki et al., 1982) and is typically apparent only 24 to several hours (38–56) *post-mortem* (Levsen et al., 2008). This poses a quality risk to the fishing and fish processing industries. The inability to identify infected fish early *post-mortem* and separate infected from uninfected fish, makes it difficult to manage processing and quality control.

The most common method for the detection of parasites and parasitic worms in fish and removal thereof, is by manual visual inspection on a light table (candling) and by manual trimming of fillets (Cheng et al., 2013; Cunningham, 2002). This method is, however, most effective for parasitic nematodes and not applicable to *K. thyrsites*, since the spores of *K. thyrsites* vary between 12.0 μm (Kabata & Whitaker, 1981) and 16.7 μm (Cunningham, 2002; Lom & Dyková, 2006) in diameter. Traditional methods for the detection of *Kudoa* infection, such as the counting of visible pseudocysts, microscopic examinations and polymerase chain reaction (PCR), are time-consuming and destructive to the sample, making these methods impractical for the fishing and fish processing industries.

Several optical or spectroscopic systems, either as stand-alone or in combination, are used for online, non-invasive and non-destructive, rapid food quality monitoring. These optical systems include, for example: near-infrared (NIR), visible (VIS), mid-infrared (MIR), Raman, nuclear magnetic resonance (NMR) spectroscopy, hyperspectral imaging (HSI) and nuclear magnetic resonance imaging (NMRI) (Cheng et al., 2013; Mathiassen et al., 2011). All of these optical systems have been applied in fish quality monitoring, including the proximate composition and lipid characteristics (Brown et al., 2012; Karlsdottir et al., 2014; Khodabux et al., 2007), bacterial spoilage (Cheng et al., 2013; Lin et al., 2006; Weeranantanaphan et al., 2011), freshness (Bøknæs et al., 2002; Ottavian et al., 2013; Reis et al., 2017; Sivertsen, Kimiya, & Heia, 2011; Uddin & Okazaki, 2004), texture analyses (Careche et al., 2002; Cheng et al., 2014; Isaksson et al., 2001) and presence of nematode parasites (Heia et al., 2007; Sivertsen et al., 2012; Sivertsen, Heia, et al., 2011; Wold et al., 2001). Due to the impracticality of using manual sorting methods to separate parasite infected fish from uninfected fish, several researchers have demonstrated the use of imaging spectroscopy as a rapid, non-destructive method separating parasite infected from uninfected fish based on the optical properties of fish muscle and parasites (Sivertsen et al., 2012; Stormo et al., 2004, 2007). Heia et al. (2007) used VIS/NIR imaging in transmission mode to identify nematode parasites on the surface of cod (*Gadus morhau*) fillets, as well as those embedded as deep as 0.8 cm into the fillet. This enabled detection of nematodes deeper into the fillet than typically possible by manual inspection. Wold et al. (2001) demonstrated the ability of multispectral imaging (400 to 1000 nm) in combination with soft independent modelling of class analogy (SIMCA) to distinguish nematode parasite infected from uninfected cod fillets.

SIMCA is a flexible, problem-dependent, multi-application use of principal component analyses (PCA)-modelling, where objects in one class or group show similar rather than identical behaviour (Wold et al., 1987). A SIMCA model is typically based on separate principal component (PC) models, one for each class or classification group. A calibration samples set is then used to build the SIMCA model by calculating PC models for the separate class groups. An ideally independent test sample set can then be subjected to the same PC model projections for classification into the separate class groups. Partial least squares discriminant analysis (PLS-DA) is another type of multivariate data analysis method that is used for classification of samples into different classes or groups based on similarity (Ballabio & Consonni, 2013; Brereton & Lloyd, 2014; Gromski et al., 2015; Grootveld, 2014; Szymańska et al., 2012). Classification of expired and non-expired vacuum packaged smoked salmon based on PLS-DA models and hyperspectral imaging spectroscopy in the short-wave NIR has been demonstrated by Ivorra et al. (2013). Ottavian et al. (2013) investigated the use of different PLS-DA models for classification of frozen–thawed fish samples independent of fish species. These authors demonstrated overall classification accuracies between 80% and 91%, depending on the PLS-DA strategies used.

The possibility of using NIR spectroscopy as a rapid, non-invasive quality control method for the detection of *Kudoa*-parasite infection before myoliquefaction is visible, especially in sardines (*Sardinops sagax ocellatus*) and Cape hake (*Merluccius capensis* and *M. paradoxus*), would be of great value to the fish processing industry. NIR spectroscopic classification of fish as either being 'parasite free' or 'parasite infected' will enable informed decision-making during quality control regarding further storage or immediate processing. The infection by *K. thyrsites* and *K. paniformis* is considered a quality risk, rather than a consumer health risk (Alvarez-Pellitero & Sitja-Bobadilla, 1993). The development of a method than can rapidly identify *Kudoa* infected fish is crucial for optimum utilization of *Kudoa* infected fish (Samaranayaka et al., 2006). Alternative utilization of *K. thyrsites* fish may include production of fish protein hydrolysates (FPH), therefore increasing its market value. However, up to date, no such rapid detection method, specifically for the

detection of *Kudoa* infection, has been reported. Although several studies have demonstrated the application of VIS/NIR (Wold et al., 2001) and hyperspectral imaging (HSI) (Sivertsen et al., 2012; Sivertsen, Heia, et al., 2011) as effective quality control methods for industrial, non-destructive detection of parasitic nematodes in fish and fish fillets, no research has been documented about the use of NIR spectroscopy for the detection of specifically *Kudoa* species. Neither has a handheld instrument been considered. The aim of this study was to investigate the use of a handheld, MicroNIR spectrophotometer, in combination with SIMCA and PLS-DA classification techniques, as a non-destructive, rapid method for the discrimination of *K. thyrsites* infected and uninfected fish species, including sardines (*Sardinops sagax ocellatus*), Cape hake (*Merluccius capensis*) and kingklip (*Genypterus capensis*).

2 | MATERIALS AND METHODS

2.1 | Samples

A total of 200 sardine, 64 Cape hake and 70 kingklip samples were supplied by the Department of Forestry, Fisheries and the Environment (DFFE, Cape Town) from several batches of pelagic and demersal surveys. All samples were kept frozen at −18°C until the presence of *Kudoa thyrsites* was determined with the use of the real-time quantitative polymerase chain reaction (qPCR) protocol as a qualitative diagnostic tool. The extraction of DNA for qPCR analysis was performed before the NIR spectra have been recorded to ensure no DNA transfer between samples and prevention of false positives.

2.2 | DNA-based identification of Kudoa thyrsites infection

DNA from composite muscle tissue samples were extracted with the use of the Qiagen DNeasy Mini Kit (Qiagen, Whitehead Scientific Pty., Ltd., Cape Town). The composite tissue samples were collected from the anterior, centre and posterior location of each fish fillet. Sterile scalpels were used for each sample to prevent DNA transfer between samples. A custom TaqMan® single nucleotide polymorphism (SNP) genotyping assay was used for the determination of the presence of *K. thyrsites* at the Centre for Proteomic and Genomic Research (CPGR, Cape Town, South Africa).

The *Kudoa thyrsites* assay mix contained two PCR primers (5′-CCTATCAACTAGTT-GGTGAGGTAGTG-3′ and 5′-TCTCC GGAATCGAACCCTGAT-3′) that flanked the region containing the SNP site specific to *K. thyrsites*. The assay mixes also contained two TaqMan® probes, one that matched the heterologous allelic variant *K. thyrsites* (5′-ACCCGTTAACACCTTG-3′), labelled with VIC (a fluorescent dye, proprietary to Life Technologies; Anonymous, 2017) and another that matched the homologous allelic variant: *K. paniformis*/*K. miniauriculata*/*K. dianae* (5′-CGTCACAACCTTG-3′), labelled with FAM (6-carboxyfluorescein). A template control (NTC) containing no DNA, a positive control and a fluorescent quencher (NRQ) were included in each assay. All TaqMan® assays were performed in 384-well plates by amplifying 20 ng DNA in a 5 μL volume PCR reaction mix containing 0.9 μM primers and 0.2 μM probes. A working master mix for each sample in each assay was prepared by mixing 0.25 μL of TaqMan® assay mix (20×), 2.5 μL TaqMan® Genotyping Master Mix (Thermo Fisher), 0.25 μL water and 2 μL of genomic DNA at 20 ng. Reactions were performed with the following protocol on the 7900HT Fast Real-Time PCR System: a pre-reading of the plate to record the background fluorescence; followed by the standard PCR cycle. The PCR cycle was performed with the following parameters: 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 1 min and post-reading of the plate. End-point fluorescence was read on the 7900HT Fast Real-Time PCR System using the SDS version 2.3 software (Applied Biosystems, USA). The TaqMan® Genotyper software 1.3 was used to generate allelic discrimination plots for the alleles: *K. thyrsites* (allele 1 – labelled with VIC) and *K. paniformis*/*K. miniauriculata*/*K. dianae* (allele 2 – labelled with FAM).

2.3 | NIR spectral acquisition

NIR absorbance spectra were collected using a MicroNIR 1700 (VIAVI Solutions, formerly JSDU, Santa Rosa, CA, USA) miniature NIR spectrophotometer. The MicroNIR instrument consists of a linear variable filter as the dispersing element, that is coupled to a linear detector array (128-pixel uncooled InGaAs photodiode array), containing a pair of integrated vacuum tungsten lamps as light source and a 16-bit analogue-to-digital converter for analogue conversion. NIR spectra were collected in the NIR region (908 to 1680 nm) of the electromagnetic spectrum at 6.2 nm intervals, resulting in 125 wavelength bands. The MicroNIR spectrophotometer was used in diffuse reflection mode and measurement parameters of integration time and number of scans were set to 14,600 μs and 50 scans. All fish samples were thawed and conditioned at room temperature for 20 min before being scanned. Three spectra were collected per fish sample at the posterior, centre and anterior anatomical locations and the average for each sample was calculated. Spectra were collected from only the flesh side (inside of the fillet) of each fish sample to capitalize on differences in reflectance values related to softer tissue (Cheng et al., 2014; Shan et al., 2018) because of myoliquefaction related to *K. thyrsites* infection.

2.4 | Multivariate data analyses

All multivariate data analyses were performed with the use of The Unscrambler software, version 10.4.1 (CAMO software, 2017, Oslo, Norway). Graphs and figures were generated with use of The Unscramble software and Microsoft Excel (Microsoft Office, 2016). Multiplicative scatter correction (MSC) was applied to the NIR

spectra to enhance spectral characteristics and to reduce undesirable sources of variation. Savitzky–Golay firts- and second-derivatives were also tested as pre-treatment techniques to investigate any improvement in classification results.

Exploratory PCA was performed on pre-treated spectra for each fish species, respectively, as well as for all fish samples combined (species independent sample set) to determine any degree of clustering between *K. thyrsites* infected and uninfected samples.

2.5 | Calibration and test sets

SIMCA and PLS-DA models for classification of *K. thyrsites* infected and uninfected fish samples were developed for sardine, Cape hake and kingklip, respectively, as well as for all the samples together (species independent sample set). To build and test the classification models, calibration and test samples sets (Table 1) were generated. This was done by randomly placing one third of the infected and uninfected samples for each respective fish species into the test set; thus, leaving two-thirds of the data set as the calibration set. In the case of the species independent sample set, it was ensured that both the calibration and test sets contained sardine, Cape hake and kingklip samples for both the infected and uninfected classes. Full cross-validation (CV) was used as validation method when classification models were developed which were then tested on the independent test sets. The number of samples used in the calibration and test sets, respectively, are summarized in Table 1.

2.6 | Generating classification models: SIMCA and PLS-DA

SIMCA was used to develop SIMCA classification models for each fish species and for the species independent data set. The statistical

TABLE 1 Number of samples in calibration and test sets for the classification classes: *Kudoa thyrsites* infected and uninfected fish samples (sardines, Cape hake and kingklip), used for the development of SIMCA and PLS-DA models and independent testing.

	Classification classes		
Sample sets	Infected fish	Uninfected fish	Total number of samples
Calibration set			
Sardines	120	14	134
Cape Hake	33	10	43
Kingklip	17	29	46
Test set			
Sardines	59	07	66
Cape Hake	17	04	21
Kingklip	09	15	24
Total number of samples	255	79	334

significance level for the SIMCA classification was set at 5%. This assumed that there was a 5% risk that a particular test sample would fall outside the class, even if it belonged to the class; while 95% of the test samples which truly belonged to the class would fall inside the class (Esbensen, 1994). Explained variance plots for the PCA class models were studied to determine the number of factors to be used for prediction of the test set samples.

Graphical interpretation of the SIMCA classification results were done with the use of Coomans plots where the transverse distances from all the test samples to the two classes were visualized. To investigate correct classification of test samples, the leverages were also studied in the distance vs. leverage plot (Si vs. Hi), also called the membership plot. The membership plot showed the limits used in the classification for both the distance to the model (the residual standard deviation) and the leverage (distance to model centre) measured for each sample. Test samples that fell inside these limits were highly likely to belong to the model or class at the chosen significance level of 5%. The model distance plot was studied to visualize the distance between the two class models, namely *K. thyrsites* infected and uninfected and to quantify whether the models were different. The variable discrimination power plots were evaluated to determine the discrimination power of each variable in the two-model comparison. A value near 1 indicated no discrimination power, while a value greater than 3 indicated good discrimination for a particular variable (Esbensen, 1994).

A PLS-DA model was developed for each fish species separately, as well as for all fish species together (species independent data set) to discriminate fish samples with *K. thyrsites* present (infected samples) from uninfected samples. PLS-DA is a classification technique used to separate different groups of samples by linking two data matrices, namely the independent variables, *X* (NIR spectra) and the dependent variables, *Y* (class membership) (Brereton & Lloyd, 2014; Gromski et al., 2015; Szymańska et al., 2012). The differences between the two groups of samples were modelled with the partial least squares (PLS) regression algorithm, but coding for class membership using the response variable −1 for members in one class and +1 for members in the second class. In this study, fish samples that were identified as *K. thyrsites* infected were placed in the 'infected' class with response variable +1, while those samples not infected were placed in the 'uninfected' class with response variable −1. The calibration set was used to run the PLS regression (with full CV). The explained variance plots were studied to determine the number of factors to be used for prediction of the test set samples. The developed model was then used to classify the samples in the test set. Test samples with predicted values ≥0 were assigned to the infected class, while test samples <0 were assigned to the uninfected class.

Classification performances of SIMCA and PLS-DA were evaluated in terms of (i) sensitivity; which represents the confidence of the class space, (ii) specificity; which is the fraction of samples not belonging to the modelled class that is correctly rejected by the model and (iii) precision; which is the ratio of the number of samples correctly accepted and the total number of samples accepted by the model (Oliveri & Downey, 2012; Szymańska et al., 2012). Calculations of these were as follows:

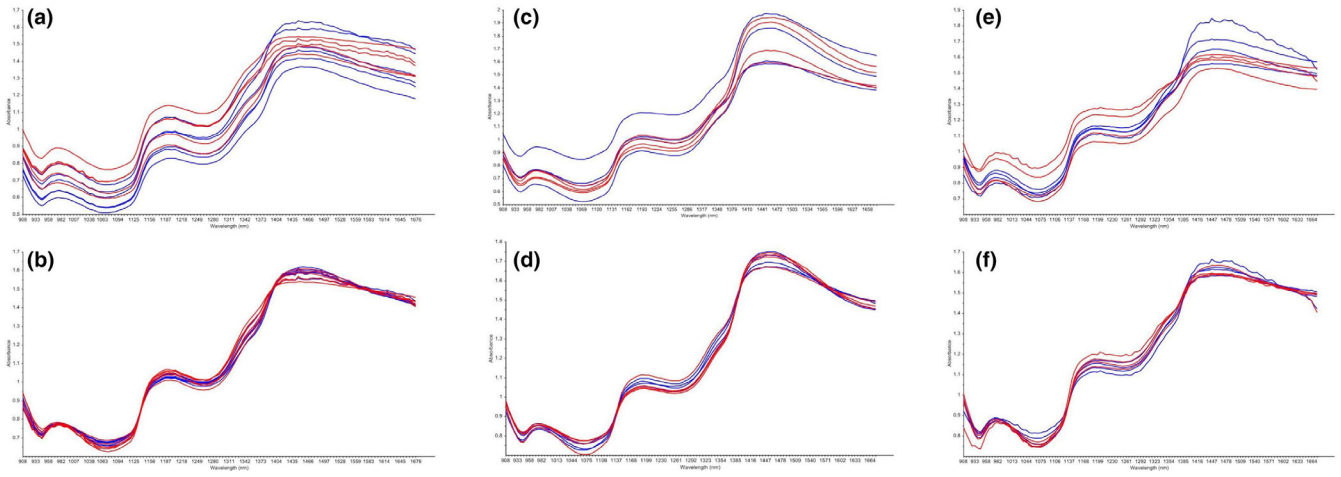

FIGURE 1 Representative raw NIR spectra for sardine (a), Cape hake (c) and kingklip (e) data sets and MSC pre-treated spectra for sardine (b), Cape hake (d) and kingklip (f) data sets, respectively. Blue lines represent *Kudoa thyrsites* infected samples and red lines represent uninfected samples.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

where (i) TP is the true positive samples correctly classified inside the class, (ii) FN is the false negative samples falling outside the classified class, (iii) FP is the false positive samples extraneous to that class but classified within the class and (iv) TN is the true negative samples classified correctly outside the class. In the case of SIMCA, test samples classified into both classes and test samples not classified in any of the two classes, were counted as misclassifications. The number of misclassifications (NMC) was calculated as the sum of FP and FN (Szymańska et al., 2012). Confusion matrices were generated for the PLS-DA classification results.

The number of samples in the two classes were unequal for all the data sets. To compensate for this, the X-matrix for the PLS model was weight centred by subtracting the average of the means of the two class groups from the column values (Brereton & Lloyd, 2014). This was done using Microsoft Excel (Microsoft Office, 2016) to investigate any improvement of the classification performance of the PLS-DA models.

3 | RESULTS AND DISCUSSION

3.1 | Quantitative polymerase chain reaction for qualitative analysis

Results from the TaqMan® qPCR assays for *Kudoa thyrsites* showed that 179 sardines (prevalence of 90%), 50 Cape hake (prevalence of 78%) and 26 kingklip (prevalence of 37%) samples were infected with *K. thyrsites*. None of the samples tested positive for the presence of allele: *K. paniformis*/*K.* miniauriculata/*K. dianae*. The prevalence of *K. thyrsites* for all 334 samples analysed was 76%. Therefore, the total 255 samples that tested positive for *K. thyrsites* infection and the 79 samples that tested negative for *K. thyrsites* infection used to generate classification models. Due to the high prevalence of *K. thyrsites* in sardines, the calibration sample set for the sardines contained 120 samples in the infected class and only 14 in the uninfected class. This resulted in the number of samples for the two classes to be unequal, which is not ideal for generating classification models based on NIR spectra (Brereton & Lloyd, 2014; Grootveld, 2014). If the one class is represented by a much larger number of samples compared to the second class, it may result in a classification model to be biased towards the class being represented by the larger number of samples. The number of samples in the two classes for the Cape hake and kingklip calibration sets were also unequal, however, the difference in number of samples was much smaller compared to that of the sardine calibration set.

3.2 | Spectral characterization and PCA

Figure 1 shows some typical absorbance spectra for the raw and MSC pre-treated NIR spectral data for representative samples from the sardine, Cape hake and kingklip data sets. Visually it was not possible to observe significant differences between the spectra for infected and uninfected samples. PCA were performed on the MSC pre-treated spectra for all fish species separately, as well as for all fish samples together, to investigate any form of clustering associated with the presence of *K. thyrsites* (Figure 2). Savitzky–Golay firts- and second-derivatives pre-treatment were also performed, but did not improve PCA models.

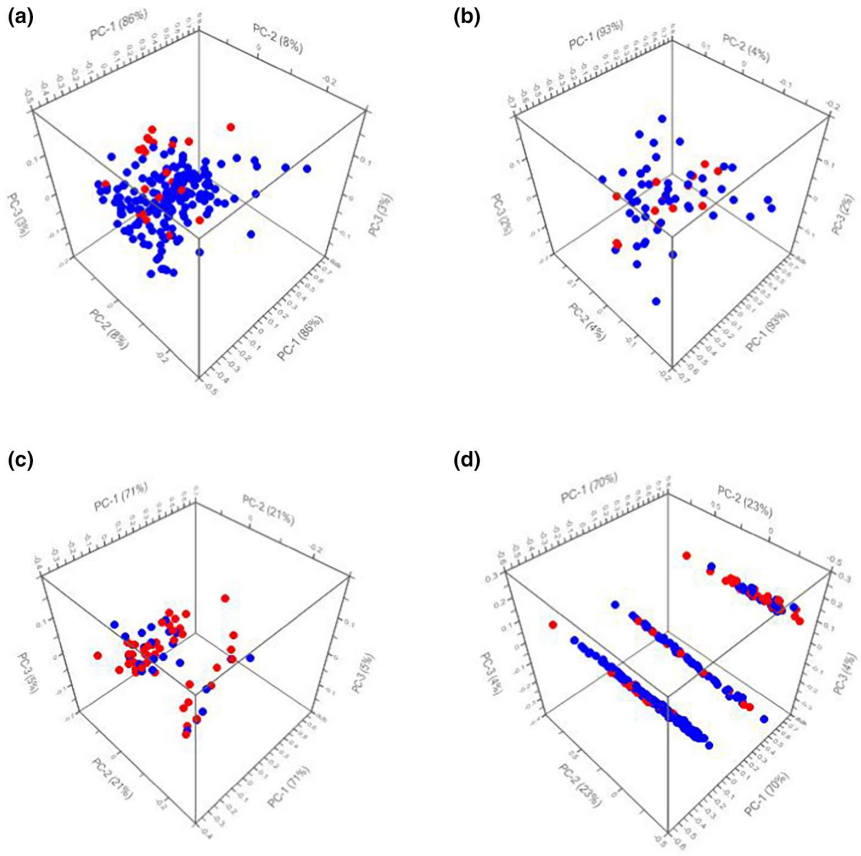

FIGURE 2 Three-dimensional (3D) PCA score plots for MSC pre-treated data, showing PC1 vs. PC2 vs. PC3 for the sardine (a), Cape hake (b), kingklip (c) and species independent (d) data sets. *Kudoa thyrsites* infected samples are represented as blue dots and uninfected samples as red dots.

The three-dimensional score plots (Figure 2a–c) showing PC1 vs. PC2 vs. PC3 for sardines (variance for PC1 was 85%, PC2 was 8% and PC3 was 3%), Cape hake (variance for PC1 was 93%, PC2 was 4% and PC3 was 2%) and kingklip (variance for PC1 was 71%, PC2 was 21% and PC3 was 5%) showed no clustering of *K. thyrsites* infected and uninfected samples. This was also the case for the species independent sample set (variance for PC1 was 70%, PC2 was 23% and PC3 was 4%). Score plots for PC4, PC5 and PC6 were also investigated, however, there was no clear clustering of infected and uninfected samples when these principal components were considered. Since there was no clustering between infected and uninfected samples, the loading plots were not investigated.

Clear clustering of the different fish species (Figure 2d) was evident, with separation between sardine, Cape hake and kingklip samples. This grouping based on species of fish was expected since the fatty acid profile, fat, protein and moisture contents differ significantly between fish species (Koubaa et al., 2011; Özogul et al., 2008; Usydus et al., 2011). O'Brien et al. (2013) demonstrated the viability of NIR spectroscopy and SIMCA to classify fish according to different species.

To further investigate possible clustering between infected and uninfected samples, the means, standard deviations and the difference in mean MSC pre-treated NIR absorbance spectra for the sardine, Cape hake, kingklip and species independent data sets were investigated. This was also done to identify specific wavelengths where differences between infected and uninfected fish samples may be evident. Visually, no spectral differences between infected and uninfected samples were observed for sardines, Cape hake, kingklip. However, for the species independent spectral data set (Figure 3) the MSC pre-treated spectra showed a difference in the mean absorbance values at 890–1380 nm between infected and uninfected samples. The infected samples had lower mean values between 890 and 1380 nm compared to the uninfected samples.

The difference in mean spectra between infected and uninfected sardine samples (Figure 4, blue line) showed an absorption band between 1218 and 1342 nm and a peak at 1390 nm. These may have been related to CH_2 (C–H stretch second overtone and 2 × C–H stretch and C–H deformation) (Osborne et al., 1993); possibly associated with fatty acids (Khodabux et al., 2007). No significant absorption band or peaks were evident in the Cape hake and kingklip difference spectra.

The functional groups of components such as proteins, water, fatty acids and carbohydrates, absorb at several NIR wavelength ranges and can overlap with the absorption of other component functional groups (Osborne et al., 1993.) This makes it difficult to identify specific absorption regions from raw spectra. When visualizing the difference in the mean MSC pre-treated NIR spectra values for infected and uninfected samples for the species independent data set (Figure 4), absorption peaks at 933 nm and 1137 nm and a band at 1206–1290 nm, were observed. The peak at 933 nm may be related to C–H stretch, third overtone of CH_2 (Osborne et al., 1993) which may be associated with fatty acids (Khodabux et al., 2007). The peak at 1137 nm and the band at 1260–1290 nm could also be associated with fatty acids (2 × C–H stretch and C–H deformation modes associated with CH_3) (Khodabux et al., 2007;

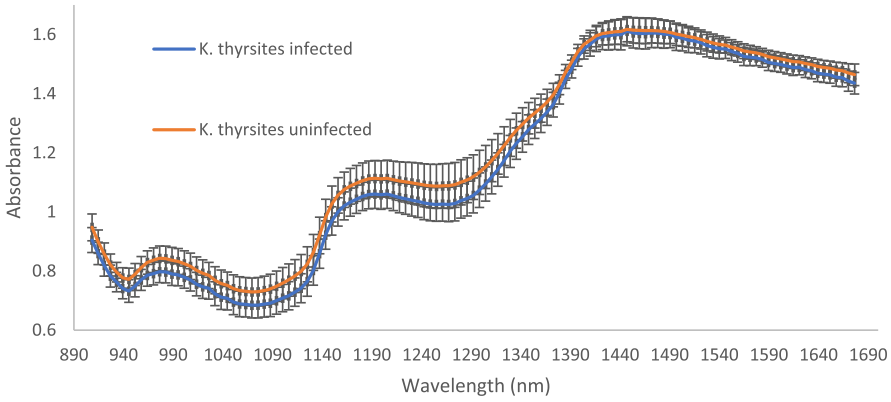

FIGURE 3 Mean MSC pre-treated NIR spectra and standard deviations for *Kudoa thyrsites* infected (blue graph) and uninfected (red graph) fish (species independent) samples.

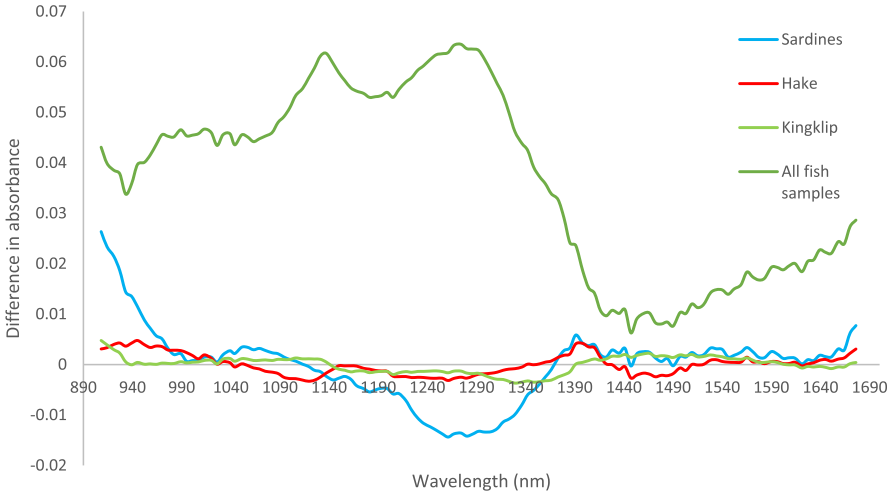

FIGURE 4 Difference in mean MSC pre-treated NIR spectra for sardines, Cape hake, kingklip and species independent data sets, respectively. Differences in mean spectra were calculated by subtracting the mean values for infected samples from mean values for uninfected samples.

Osborne et al., 1993). Absorption peaks related to fatty acids are mainly attributed to the C–H and CH_2 vibrations with the first overtone of C–H stretch in the range of 1718–1760 nm and the second overtone in the range 1100–1390 nm. In this study, it was expected to observe differences in NIR spectra between infected and uninfected samples at wavelengths associated with water, proteins, peptides and amino acids because of the extensive protein breakdown associated with *K. thyrsites* infection (Funk et al., 2008; Langdon, 1991; Samaranayaka et al., 2006; Tsuyuki et al., 1982). Extensive protein hydrolyses because of the activity of cysteine cathepsin L proteases excreted by *K. thyrsites* (Funk et al., 2008; Samaranayaka et al., 2006; Tsuyuki et al., 1982) result in release of moisture (Ofstad et al., 1996) and peptide fragments. In general, N-H stretch first overtone of $CONH_2$ of proteins and amino acids are typically visible at 1430 nm (Osborne et al., 1993) and O-H first overtone associated with water is typically visible at 1440–1450 nm. Proteins and peptides have broad, strong bands that are difficult to differentiate due to considerable overlap of the bands. NIR spectra of fish may be dominated by absorption peaks attributed to protein at 1510 nm, 1700 nm and 1738 nm (Uddin & Okazaki, 2004).

Raw fish muscle typically contains between 12% and 22% protein (Foline et al., 2011; Gokoglu et al., 2004; Koubaa et al., 2011; Usydus et al., 2011; Weber et al., 2008), while lipids and water together make up about 80% of fish muscle (Ofstad et al., 1996). The moisture content of fish muscle ranges from 65% to 80% (Bulla et al., 2011; Gokoglu et al., 2004; Roncarati et al., 2012) as influenced by catching season (Boran et al., 2008). Fish muscle has a very low carbohydrate content, ranging between 0.2% to 1.5% (Françoise, 2010). The observed differences in difference spectra for the fish species independent dataset could be due to the changes in moisture content, fatty acids and protein structure between the difference fish species. These changes probably occurred during frozen storage of the samples. Atanassova et al. (2024) demonstrated clear differences in NIR absorption patterns, mainly at the O–H and N–H bands, between fresh and frozen–thawed carp (*Cyprinus carpio*) samples. Similar results were obtained by Cheng et al. (2015) for reflectance values of fresh, cold-stored and frozen–thawed grass carp fillets.

Differences in NIR spectra associated with proteins and water were not evident between infected and uninfected fish samples. It is hypothesized that structural changes due to prolonged frozen storage of the fish samples used in this study reduced or masked the effects of the *K. thyrsites* associated proteolytic activity (Zhou & Li-Chan, 2009).

3.3 | Prediction ability of SIMCA and PLS-DA models

Comparing the use of Savitzky–Golay firts- and second-derivatives as pre-treatment techniques with that of MSC, no improvement in

the performance of the generated SIMCA and PLS-DA classification models were obtained. Cheng et al. (2015) also reported that spectral pre-processing techniques such as standard normal variate and firts- and second-derivatives did not improve the performance of classification models, including SIMCA and PLS-DA, for discrimination between fresh and frozen–thawed fish. On the other hand, a study by Shan et al. (2018) illustrated that the effects of spectral pre-processing techniques depended on the type of spectral noise generated from whether the spectra were obtained from the skin-side or from the flesh-side of fish. These authors found that PLS models developed using spectra obtained from the fish-flesh and pre-processing by the second derivative method, resulted in the best prediction models for storage time of fresh and frozen–thawed fish fillets; while MSC pre-processing method performed better with spectra from the skin-side of the fish.

Pertaining to the SIMCA classification, the number of factors used for the sardine test set were two for the PCA infected model and three for the PCA uninfected model. For Cape hake, two factors were used for the PCA infected model and one factor for the uninfected PCA model. For kingklip and the species independent test sets, three factors were used for both the infected and uninfected PCA models. Figure 5a,b shows, as examples, the Coomans plots for the results of the SIMCA classification for the sardine and the species independent test sets, respectively. In both the sardine and species independent SIMCA classification models, there were poor separation between the infected and uninfected calibration sample sets (Figure 5). This was expected, as a clear separation of the two classes was not seen in the overall PCA models of the calibration sample sets. Bächle et al. (2012) reported similar challenges with poor classification ability of SIMCA when clusters of groups were close to each other or overlapped in the PCA score plots.

Classifications of the sardine samples in the test sets were poor, since samples mostly fell outside both classes (Figure 5a for sardine) or inside both classes (Figure 5b for species independent set). Similar results were obtained from the Coomans plots for SIMCA classification of the Cape hake and kingklip test sample sets. The number of samples in both the Cape hake and kingklip total data sets were low (total samples of 64 and 70, respectively). Theoretically the number of samples in a data set should be more than 100, or 5 times the number of *X*-variables (Grootveld, 2014).

Classification results for the SIMCA models for sardines, Cape hake, kingklip and the species independent test sample sets are given in terms of the NMC (Table 2), which is the sum of false positive (FP) and false negative (FN) (Szymańska et al., 2012). Samples classified as belonging to both classes were counted as false positives (FP). Samples not classified at all were counted as FN. The number of true positives (TP) and true negatives (TN) are also indicated in Table 2.

None of the SIMCA models showed acceptable prediction ability as more than 60% of the samples in the test sets were misclassified in all the models. The species independent SIMCA classification model classified only 8% of the test set samples correctly. Seven infected samples were correctly classified as infected (TP), while two uninfected samples were correctly classified as uninfected (TN). The total NMC was 102, where 98 samples were classified into both groups, three samples were not classified at all and one infected sample was wrongly classified as uninfected (FN). SIMCA models are based on PC's (Oliveri & Downey, 2012). There was no clear clustering of infected from uninfected samples in the PCA models for sardine, Cape hake, kingklip and the species independent data sets (Figure 2). This may therefore explain the low prediction ability of the SIMCA models.

Seven factors were used for the sardine, Cape hake and kingklip PLS-DA models; while for the species independent PLS-DA model, five factors were used. Classification results for the PLS-DA models for the sardine, Cape hake, kingklip and the species independent test sample sets are summarized as confusion matrixes in Table 3. In the case of the sardines, all test samples were predicted as infected, with the NMC calculated as 7 (all being FP). Similar classification results were obtained in the case of the Cape hake test sample set. All the hake test samples ($n=21$) were classified as being infected, with 4 false positives and a NMC of 4. For both the sardine and Cape hake PLS-DA models, the sensitivity was calculated as 1, while the specificity was calculated as zero (since TN = 0 in both cases). The precision was calculated as 0.89 for sardines and 0.81 for Cape hake. These values, however, do not indicate acceptable classification ability of the models, since none of the uninfected test samples were correctly classified as uninfected.

The kingklip PLS-DA model (Table 3) gave a NMC of 12 (FP + FN), where three samples were wrongly classified as infected (FP) and nine samples were wrongly classified as uninfected (FN). For the kingklip PLS-DA model, sensitivity was calculated as zero (due to TP = 0), specificity as 0.80 and precision as zero (since TP = 0).

The species independent PLS-DA model correctly classified 91% (77 of 85 samples) of the infected test samples into the infected class and 58% (15 of 26 samples) of the uninfected samples into the uninfected class (Table 3). The sensitivity of this PLS-DA model was calculated as 0.89, the specificity as 0.58 and the precision as 0.87. Although the calculated sensitivity and precision values seem acceptable, the correctly classified uninfected samples were all kingklip samples (TN = 15), while none of the uninfected sardine and Cape hake test samples were correctly classified as uninfected (FP = 11). The 77 (TP) samples correctly classified as being infected were all sardine and Cape hake samples, while none of the infected kingklip test samples were correctly classified as infected. It seemed though as if the species independent PLS-DA model distinguished between the kingklip samples and the sardine and Cape hake samples. Although a species independent classification model would be ideal, it seems as though species-specific models might be more useful to distinguish between *K. thyrsites* infected and uninfected fish samples. It also seemed as though the PLS-DA models (species dependent and independent) classified most of the samples in the test sets as infected, thus favouring the class with the larger number of representative samples (Brereton & Lloyd, 2014; Grootveld, 2014). It may be argued that greater importance was given to the infected class since the number of samples in the different classes was unbalanced, where the infected class had a much higher number of samples

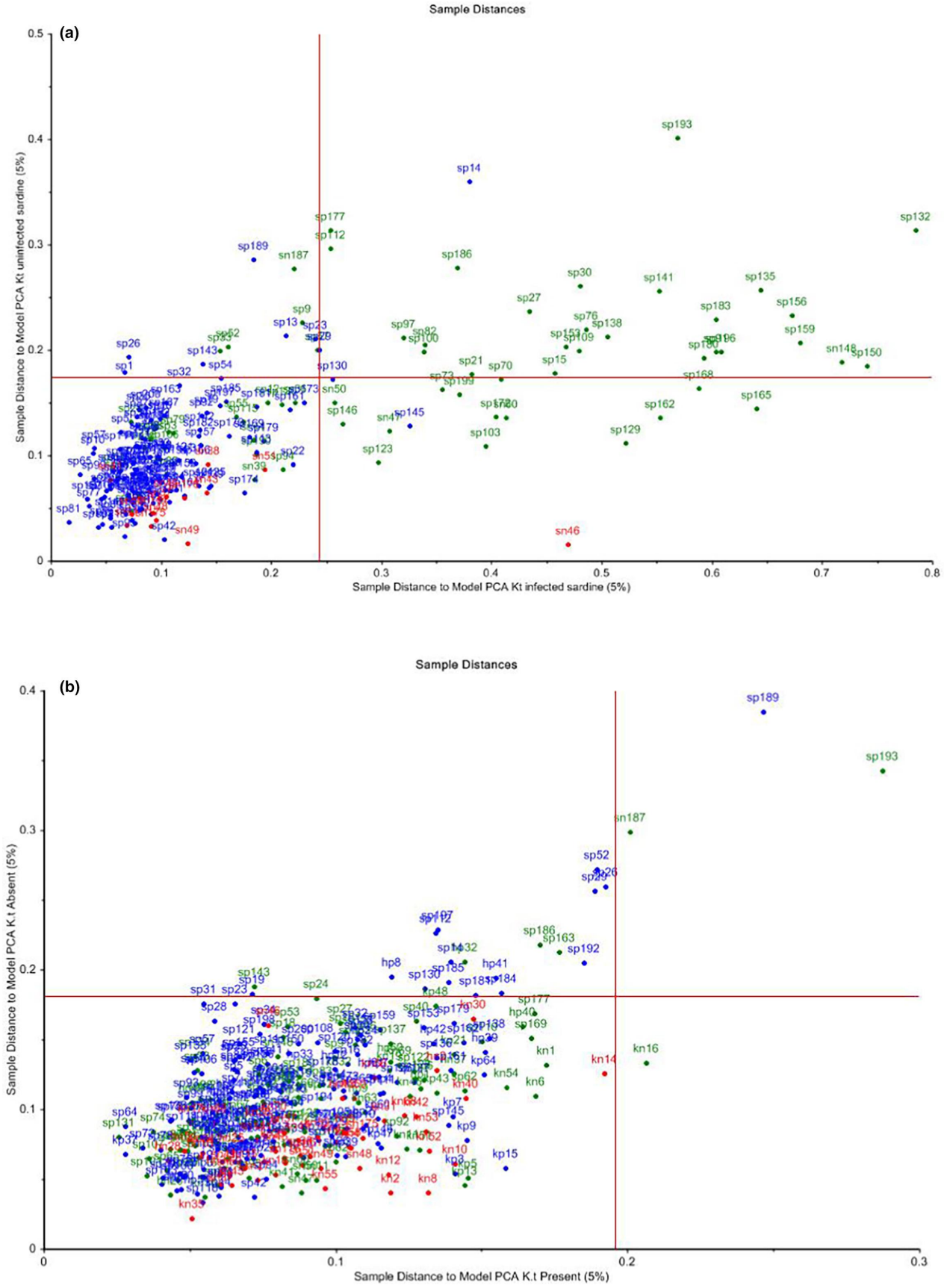

FIGURE 5 Coomans plots for SIMCA classification of the (a) sardine and (b) species independent test sample sets. Test samples are indicated in green, calibration samples for the *Kudoa thyrsites* infected class are indicated in blue; while the samples for the uninfected calibration class are indicated in red.

TABLE 2 Number of misclassifications (NMC) for SIMCA models for sardines, Cape hake, kingklip and species independent test sample sets, respectively.

SIMCA models	NMC	Total *n*	TP	TN
Sardines	61	66	5	0
Cape hake	18	21	3	0
Kingklip	23	24	1	0
Species-independent data set	102	111	7	2

Note: NMC – number of misclassifications = FP + FN.

Abbreviations: *n*, number of samples; TN, true negatives; TP, true positives.

TABLE 3 Confusion matrixes for the sardine, Cape hake, kingklip and species-independent PLS-DA models, respectively.

	Predicted uninfected	Predicted infected	Total *n*
Sardines:			
Test set sample number, *n* = 66			
Actual uninfected	0 (=TN)	7 (=FP)	7
Actual infected	0 (=FN)	59 (=TP)	59
Total *n*	0	66	66
Cape hake:			
Test set sample number, *n* = 21			
Actual uninfected	0 (=TN)	4 (=FP)	4
Actual infected	0 (=FN)	17 (=TP)	17
Total *n*	0	21	21
Kingklip:			
Test set sample number, *n* = 24			
Actual uninfected	12 (=TN)	3 (=FP)	15
Actual infected	9 (=FN)	0 (=TP)	9
Total *n*	21	3	24
Species independent sample set:			
Test set sample number, *n* = 111			
Actual uninfected	15 (=TN)	11 (=FP)	26
Actual infected	8 (=FN)	77 (=TP)	85
Total *n*	23	88	111

Abbreviations: FN, false negative; FP, false positive; *n*, number of samples; TN, true negative; TP, true positive.

than the uninfected class. When dealing with unbalanced classes, it is recommended to perform weight centring on the X-matrix for the PLS model by subtracting the average of the means of the two class groups from the column values (Brereton & Lloyd, 2014). To compensate for the unbalanced classes in this study, this was done for all the PLS-DA models. This, however, did not improve the classification ability of any of the PLS-DA models.

Similar to the SIMCA classification models, the PLS-DA models did not show acceptable classification results. Classification of samples into both classes is not seen with the use of the PLS-DA model since the predicted values for the *Y*-variable of the test samples are discriminated based on the threshold of zero (0). Predicted samples are simply classified either as belonging to the one class (predicted $y \geq 0$) or the other class (predicted $y < 0$). However, PLS-DA does not consider within-class variability (Brereton & Lloyd, 2014), making it difficult to assign samples with values close to zero (with 1 as the level of the infected class and −1 as the level of the uninfected class) into the correct class. To evaluate the performance of the PLS-DA models in this study, all test set samples with predicted *y*-values ≥0 were classified as infected and those samples <0 were classified as uninfected.

Acquisition of NIR spectra of fresh fish samples, not previously frozen, as early as 24 h *post-mortem* is strongly suggested. In a study by Atanassova et al. (2024), that investigated the use of NIR spectroscopy with SIMCA and PLS-DA to discriminate between fresh and frozen–thawed fish, it was shown that the accuracy of the models to correctly classify frozen–thawed samples was better for samples that were treated with a second freeze–thaw cycle. This was attributed to the extensive damage to muscle fibres after a second cycle of freezing and thawing, compared to samples that were freeze-thawed only once, resulting in larger structural differences between fresh and frozen–thawed fish samples. Textural changes (myoliquefaction) in the muscle tissue of fresh sardine samples, which have not been frozen (Figure 6), infected with *K. thyrsites* were observed as early as 24 h *post-mortem*. However, since some samples were frozen before aquations of NIR spectra, it is possible that any textural or chemical differences associated with the parasite between infected and noninfected samples were obscured by the effect of the frozen-thaw process.

The extent of myoliquefaction as were observed in the sardine samples, were not evident in the Cape hake and kingklip samples, however, black pseudocysts were visible in some of the Cape hake samples. Textural changes such as gaping and chemical (accumulation of melanin and formation of black pseudocysts) changes associated with *K. thyrsites* infection have been demonstrated for hake species (Kabata & Whitaker, 1981; Tsuyuki et al., 1982). In the case of Cape hake species, where white, light brown and/or black pseudocysts may be evident due to host specific response towards *K. thyrsites* (Morado & Sparks, 1986), it is suggested to investigate the use of imaging spectroscopy (Heia et al., 2007; Sivertsen et al., 2012; Sivertsen, Heia, et al., 2011; Stormo et al., 2004, 2007) in an attempt to differentiate between *K. thyrsites* infected and uninfected Cape hake samples based on optical properties of fish muscle and pseudocysts. Wold et al. (2001) demonstrated the use of multispectral imaging and SIMCA based on pixels to discriminate between different coloured parasitic worms as well as the depth within the fish fillet where the worms were located. The benefit of using multispectral imaging is that pixels associated with the parasite are used to develop the SIMCA models.

4 | CONCLUSIONS

In this study, NIR spectroscopy, in combination with SIMCA and PLS-DA, was unable to satisfactory distinguish between

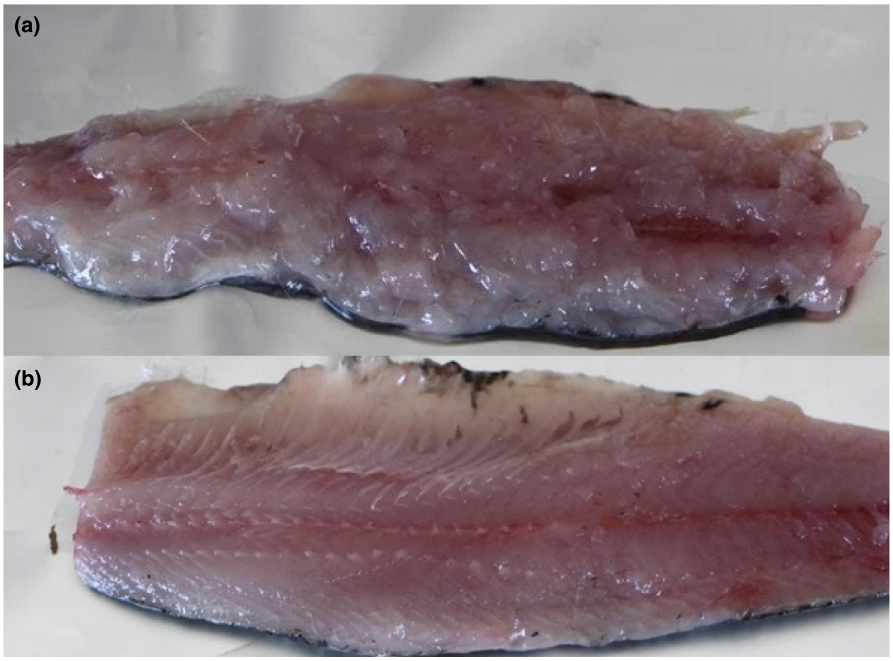

FIGURE 6 Kudoa thyrsites infected fresh sardine samples at 24 h post-mortem. (a) Extensively myoliquefied sardine samples. (b) sardine sample not yet showing myoliquefaction at 24 h post-mortem.

K. thyrsites infected and uninfected fish samples. The fish samples used for this study were frozen for up to 8 months before being analysed. Structural and chemical changes due to frozen storage may have masked the chemical changes associated with *K. thyrsites* infection, thus, differences in NIR spectra and PCA were not evident between infected and uninfected samples. Further species dependent studies are recommended. In the case of sardine, it is suggested to investigate the use of NIR spectroscopy and texture analyses, in combination with the level of *K. thyrsites* infection (namely the number of *Kudoa* spores per gram of muscle tissue), to distinguish between infected samples with the potential of developing extensive myoliquefaction and those that may still be firm and suitable for processing. It is recommended to collect NIR spectra of samples earlier than 24 h *post-mortem* and of prior to freezing. This can be done by taking the portable, hand-held, MicroNIR spectrophotometer onto fishing vessels. NIR spectra of the same samples, in addition to texture and prevalence analyses, should then be acquired again at 24 h *post-mortem* to study differences in textural properties between early *post-mortem* and 24 h *post-mortem* samples. *Kudoa thyrsites* infected fish which may not become extensively myoliquefied can then be separated and used for further processing. Infected fish which may develop extensive myoliquefaction and not suitable for processing, may be channelled to alternative processing, such as fish meal and/or fish oil manufacturing. However, since the prevalence of *K. thyrsites* in sardine is very high, it will be a challenge to analyse enough sardine samples to obtain a calibration set with two classes of about 50 samples each. A quantity of at least 500 sardine samples may be required to obtain 50 uninfected samples.

For Cape hake, it is suggested to investigate the use of imaging spectroscopy to differentiate between *K. thyrsites* infected and uninfected samples based on the presence of *K. thyrsites* associated pseudocysts. Differences in optical spectra between pseudocysts and fish muscle tissue may be significant enough to generate classification models. It is further recommended to ensure equal number of infected and uninfected samples in the data sets, although this may require analysis of large numbers of fish samples over an extended period.

AUTHOR CONTRIBUTIONS

S. Henning: Conceptualization; investigation; funding acquisition; writing – original draft; resources; methodology; project administration; writing – review and editing.

ACKNOWLEDGEMENTS

The study was supported by the South African National Research Funds (NRF), under the Thuthuka funding instrument TTK13062720022.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

S. Henning https://orcid.org/0000-0002-3624-4467

REFERENCES

Alvarez-Pellitero, P., & Sitja-Bobadilla, A. (1993). Pathology of myxosporea in marine fish culture. *Diseases of Aquatic Organisms*, *17*, 229–238. https://doi.org/10.3354/dao017229

Atanassova, S., Yorgov, D., Stratev, D., Veleva, P., & Stoyanchev, T. (2024). Near-infrared spectroscopy for rapid differentiation of fresh and frozen-thawed common carp (*Cyprinus carpio*). *Sensors*, *24*, 3620. https://doi.org/10.3390/s24113620

Bächle, H., Zimmer, B., & Wegener, G. (2012). Classification of thermally modified wood by FT-NIR spectroscopy and SIMCA. *Wood Science and Technology*, *46*, 1181–1192.

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, *5*, 3790–3798.

Bøknæs, N., Jensen, K. N., Andersen, C. M., & Martens, H. (2002). Freshness assessment of thawed and chilled cod fillets packed in modified atmosphere using near-infrared spectroscopy. *LWT – Food Science and Technology*, *35*(7), 628–634.

Boran, G., Boran, M., & Karaçam, H. (2008). Seasonal changes in proximate composition of anchovy and storage stability of anchovy oil. *Journal of Food Quality*, *31*, 503–513.

Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. *Journal of Chemometrics*, *28*, 213–225.

Brown, M. R., Kube, P. D., O'Connor, S., Cunningham, M., & King, H. (2012). Application of near infrared reflectance spectroscopy for the rapid chemical analysis of Sydney rock oyster (*Saccostrea glomerate*) and Pacific oyster (*Crassostrea gigas*). *Journal of Shellfish Research*, *31*(4), 1051–1060.

Bulla, M. K., Simionato, J. I., Matsushita, M., Coró, F. A. G., Shimokomaki, M., Visentainer, J. V., & de Souza, N. E. (2011). Proximate composition and fatty acid profile of raw and roasted salt-dried sardines (*Sardinella brasiliensis*). *Food and Nutrition Sciences*, *2*, 440–443.

Careche, M., García, M. L., Herrero, A., Solas, M. T., & Carmona, P. (2002). Structural properties of aggregates from frozen stored hake muscle proteins. *Journal of Food Science*, *67*(8), 2827–2832.

Cheng, J.-H., Dai, Q., Sun, D.-W., Zeng, X.-A., Liu, D., & Pu, H.-B. (2013). Applications of non-destructive spectroscopic techniques for fish quality and safety evaluation and inspection. *Trends in Food Science and Technology*, *34*, 18–31.

Cheng, J.-H., Sun, D.-W., Han, Z., & Zeng, X.-A. (2014). Texture and structure measurements and analyses for evaluation of fish and fillet freshness quality: A review. *Comprehensive Reviews in Food Science and Food Safety*, *13*, 52–61.

Cheng, J.-H., Sun, D.-W., Pu, H.-B., Chen, X., Liu, Y., Zhang, H., & Li, J.-L. (2015). Integration of classifiers analysis and hyperspectral imaging for rapid discrimination of fresh from cold-stored and frozen-thawed fish fillets. *Journal of Food Engineering*, *161*, 33–39. https://doi.org/10.1016/j.jfoodeng.2015.03.011

Cunningham, C. O. (2002). Molecular diagnosis of fish and shellfish diseases: Present status and potential use in disease control. *Aquaculture*, *206*, 19–55. https://doi.org/10.1016/S0044-8486(01)00864-X

Dawson-Coates, J.A., Chase, J.C., Funk, V., Booy, M.H, Haines, L.R., Falkenberg, C.L., Whitaker, D.J., Olafson, R.W. & Pearson, T.W. (2003). The relationship between flesh quality and numbers of *Kudoa thyrsites* plasmodia and spores in farmed Atlantic salmon, *Salmo salar* L. *Journal of Fish Diseases*, *26*, 451–459. https://doi.org/10.1046/j.1365-2761.2003.00477.x

Esbensen, K. H. (1994). *Multivariate data analysis – In practice. An introduction to multivariate data analysis and experimental design* (5th ed., pp. 19, 335–104, 360). CAMO.

Foline, O. F., Rachael, A. M., Iyabo, B. E., & Fidelis, A. E. (2011). Proximate composition of catfish (*Clarias gariepinus*) smoked in Nigerian stored products research institute (NSPRI): Developed kiln. *International Journal of Fisheries and Aquaculture*, *3*(5), 95–97.

Françoise, L. (2010). Occurrence and role of lactic acid bacteria in seafood products. A review. *Food Microbiology*, *27*, 698–709.

Funk, V. A., Olafson, R. W., Raap, M., Smith, D., Aitken, L., Haddow, J. D., Wang, D., & Dawson-Coates, J. A. (2008). Identification, characterization and deduced amino acid sequence of the dominant protease from *Kudoa paniformis* and *K. thyrsites*: A unique cytoplasmic cysteine protease. Comparative biochemistry and physiology. *Part B: Biochemistry and Molecular Biology*, *149*, 477–489. https://doi.org/10.1016/j.cbpb.2007.11.011

Gilchrist, J. D. F. (1924). A protozoal parasite *Chloromyxum thyrsites* sp.n. of the Cape sea-fish the 'snoek' (*Thyrsites atun*, EUPHR). *Transactions of the Royal Society of South Africa*, *11*, 263–273. https://doi.org/10.1080/00359192309519587

Gokoglu, N., Yerlikaya, P., & Cengiz, E. (2004). Effects of cooking methods on the proximate composition and mineral contents of rainbow trout (*Oncorhynchus mykiss*). *Food Chemistry*, *84*, 19–22.

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis – A marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, *879*, 10–23.

Grootveld, M. (2014). Introduction to the applications of chemometric techniques in 'omics' research: Common pitfalls, misconceptions and 'rights and wrongs'. In D. Anderson & A. Groorveld (Eds.), *Metabolic profiling: Disease and xenobiotics* (pp. 1–33). Royal Society of Chemistry.

Heia, K., Sivertsen, H. A., Stormo, S. K., Elvevoll, E., Wold, J. P., & Nilsen, H. (2007). Detection of nematodes in cod (*Gadus morhau*) fillets by imaging spectroscopy. *Journal of Food Science*, *72*(1), E11–E15.

Anonymous. (2017). *VIC Fluorescent dye (ABI)*. Retrieved June 02, 2017, from http://www.genelink.com/newsite/products/mod_detail.asp?modid=63.

Huss, H. H. (1994). In H. H. Huss (Ed.), *Assurance of seafood quality* (Vol. *334*). FAO Fisheries Technical Paper.

Isaksson, T., Swensen, L. P., Taylor, R. G., Fjæra, S. O., & Skjervold, P. O. (2001). Non-destructive texture analysis of farmed Atlantic salmon using visual/near-infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, *82*, 53–60.

Ivorra, E., Girón, J., Sánchez, J., Verdú, S., Barat, J. M., & Grau, R. (2013). Detection of expired vacuum-packed smoked salmon based on PLS-DA method using hyperspectral images. *Journal of Food Engineering*, *117*, 342–349.

Kabata, Z., & Whitaker, D. J. (1981). Two species of *Kudoa* (Myxosporea, Multivalvulida) parasite in the flesh of *Merluccius productus* (Ayres, 1855) (Pisces, Teleostei) in the Canadian Pacific. *Canadian Journal of Zoology*, *59*, 2085–2091. https://doi.org/10.1139/z81-285

Karlsdottir, M. G., Arason, S., Kristinsson, H. G., & Sveinsdottir, K. (2014). The application of near infrared spectroscopy to study lipid characteristics and deterioration of frozen lean fish muscles. *Food Chemistry*, *159*, 420–427.

Khodabux, K., L'Omelette, M. S. S., Jhaumeer-Laulloo, S., Ramasami, P., & Rondeau, P. (2007). Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. *Food Chemistry*, *102*(3), 669–675.

Koubaa, A., Abdelmouleh, A., Bouain, A., & Mihoubi, N. B. (2011). Experimental and statistical investigations of the global chemical composition of sixtrawling fish of the Gulf of Gabès (Mediterranean Sea). *Journal of the Marine Biological Association of the United Kingdom*, *91*(8), 1689–1695. https://doi.org/10.1017/S0025315410000640

Langdon, J. S. (1991). Myoliquefaction *post-mortem* ('milky flesh') due to *Kudoa thyrsites* (Gilchrist) (Myxosporea: Multivalvulida) in mahi mahi, *Coryphaena hippurus* L. *Journal of Fish Diseases*, *14*, 45–54. https://doi.org/10.1111/j.1365-2761.1991.tb00575.x

Levsen, A., Jørgensen, A., & Mo, T. A. (2008). Occurence of post-mortem myoliquefactive kudoosis in Atlantic mackerel, *Scomber scombrus* L., from the North Sea. *Journal of Fish Diseases*, *31*, 601–611. https://doi.org/10.1111/j.13652761.2008.00937.x

Lin, M., Mousavi, M., Al-Holy, M., Cavinato, A. G., & Rasco, B. A. (2006). Rapid near-infrared spectroscopic method for the detection of spoilage in rainbow trout (*Oncorhychus mykiss*) fillet. *Journal of Food Science*, *71*(1), S18–S23.

Lom, J., & Dyková, I. (2006). Myxozoan genera: Definition and notes on taxonomy, lifecycle terminology and pathogenic species. *Folia Parasitologica*, *53*, 1–36.

Mathiassen, J. R., Misimi, E., Bondø, M., Veliyulin, E., & Østvik, S. O. (2011). Trends in application of imaging technologies to inspection

of fish and fish products. *Trends in Food Science and Technology*, *22*, 257–275.

Morado, J. F., & Sparks, A. (1986). Observations on the host-parasite relations of the Pacific whiting, *Merluccius productus* (Ayres), and two myxosporean parasites, *Kudoa thyrsites* (Gilchrist, 1924) and *K. paniformis* Kabata & Whitaker, 1981. *Journal of Fish Diseases*, *9*, 445–455. https://doi.org/10.1111/j.1365-2761.1986.tb01038.x

O'Brien, N., Hulse, C. A., Pfeifer, F., & Siesler, H. W. (2013). Near infrared spectroscopy authentication of seafood. *Journal of Near Infrared Spectroscopy*, *21*, 299–305. https://doi.org/10.1255/jnirs.1063

Ofstad, R., Egelandsdal, B., Kidman, S., Myklebust, R., Olsen, R. L., & Hermansson, A.-M. (1996). Liquid loss as effected by *post-mortem* ultrastructural changes in fish muscle: Cod (*Gadus morhau* L) and Salmon (*Salmo salar*). *Journal of the Science of Food and Agriculture*, *71*, 301–312.

Oliveri, P., & Downey, G. (2012). Multivariate class modelling for the verification of food-authenticity claims. *Trends in Analytical Chemistry*, *35*, 74–86.

Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman Scientific and Technical.

Ottavian, M., Fasolato, L., Facco, P., & Barolo, M. (2013). Foodstuff authentication from spectral data: Toward a species independent discrimination between fresh and frozen-thawed fish samples. *Journal of Food Engineering*, *119*, 765–775.

Özogul, Y., Özogul, F., Çi˙Çek, E., Polat, A., & Kuley, E. (2008). Fat content and fatty acid composition of 34 marine water fish species from the Mediterranean Sea. *International Journal of Food Sciences and Nutrition*, *60*(6), 464–475.

Reis, M. M., Martínez, E., Saitua, E., Rodríguez, R., Pérez, I., & Olabarrieta, I. (2017). Non-invasive differentiation between fresh and frozen/thawed tuna fillets using near infrared spectroscopy (vis-NIRS). *LWT – Food Science and Technology*, *78*, 129–137.

Roncarati, A., Brambilla, G., Maluzzi, A., Iamiceli, A. L., Fanelli, R., Moret, I., Ubaldi, A., Miniero, R., Sirri, F., Melotti, P., & di Domenico, A. (2012). Fatty acid profile and proximate composition of fillets from *Engraulis encrasicholus*, *Mullus barbatus*, *Merluccius merluccius* and *Sarda sarda* caught in Tyrrhenian, Adriatic and Ionian seas. *Journal of Applied Ichthyology*, *28*, 545–552.

Samaranayaka, A. G. P., Ho, T. C. W., & Li-Chan, E. C. Y. (2006). Correlation of *Kudoa* spore counts with proteolytic activity and texture of fish mince from Pacific hake (*Merluccius productus*). *Journal of Aquatic Food Product Technology*, *15*(4), 75–93.

Shan, J., Wang, X., Russel, M., Zhao, J., & Zhang, Y. (2018). Comparison of fish morphology for fresh and frozen-thawed crucian carp quality assessment by hyperspectral imaging technology. *Food Analytical Methods*, *11*, 1701–1710. https://doi.org/10.1007/s12161-018-1158-5

Sivertsen, A. H., Heia, K., Hindberg, K., & Godtliebsen, F. (2012). Automatic nematode detection in cod fillets (*Gadus morhau* L.) by hyperspectral imaging. *Journal of Food Engineering*, *111*, 675–681.

Sivertsen, A. H., Heia, K., Stormo, S., Elvevoll, E., & Nilsen, H. (2011). Automatic nematode detection in cod fillets (*Gadus morhau*) by transillumination hyperspectral imaging. *Journal of Food Science*, *76*(1), S77–S83.

Sivertsen, A. H., Kimiya, T., & Heia, K. (2011). Automatic freshness assessment of cod (*Gadus morhua*) fillets by Vis/Nir spectroscopy. *Journal of Food Engineering*, *103*, 317–323.

Stormo, S. K., Ernstsen, A., Nilsen, H., Heia, K., Sivertsen, A. H., & Elvevoll, E. (2004). Compounds of parasitic roundworm absorbing in the visible region: Target molecules for detection of roundworm in Atlantic cod. *Journal of Food Protection*, *67*(7), 1522–1525.

Stormo, S. K., Sivertsen, A. H., Heia, K., Nilsen, H., & Elvevoll, E. (2007). Effects of single wavelength selection for anisakid roundworm larvae detection through multispectral imaging. *Journal of Food Protection*, *70*(8), 1890–1895.

Szymańska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, *8*, S3–S16.

Tsuyuki, H., Williscroft, S. N., Kabata, Z., & Whitaker, D. J. (1982). *The relationship between acid and neutral protease activities and the incidence of soft cooked texture in the muscle tissue of Pacific hake Merluccius productus infected with* Kudoa paniformis *and/or* K. thyrsites, *and held for varying times under different prefreeze chilled storage conditions. Canadian technical report of fisheries and aquatic sciences No. 1130*. Ottawa: Department of Fisheries and Oceans; 1982. p. 1–39.

Uddin, M., & Okazaki, E. (2004). Classification of fresh and frozen-thawed fish by near-infrared spectroscopy. *Journal of Food Science*, *69*(8), C665–C668.

Usydus, Z., Szlinder-Richert, J., Adamczyk, M., & Szatkowska, U. (2011). Marine and farmed fish in the polish market: Comparison of the nutritional value. *Food Chemistry*, *126*, 78–84.

Webb, S. C. (1990). Pap pilchards and protozoa. *South African Shipping News & Fishing Industry Review*, *45*(4), 35.

Weber, J., Bochi, V. C., Ribeiro, C. P., de Victório, A. M., & Emanuelli, T. (2008). Effect of different cooking methods on the oxidation, proximate and fatty acid composition of silver catfish (*Rhamdia quelen*) fillets. *Food Chemistry*, *106*, 140–146.

Weeranantanaphan, J., Downey, G., Allen, P., & Sun, D.-W. (2011). A review of near infrared spectroscopy in muscle food analysis: 2005-2010. *Journal of Near Infrared Spectroscopy*, *19*, 61–104.

Wold, J. P., Westad, F., & Heia, K. (2001). Detection of parasites in cod fillets by using SIMCA classification in multispectral images in the visible and NIR region. *Applied Spectroscopy*, *55*(8), 1025–1034.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*, 37–52.

Zhou, L. S., & Li-Chan, E. C. Y. (2009). Effects of *Kudoa* spores, endogenous protease activity and frozen storage on cooked texture of minced Pacific hake (*Merluccius productus*). *Food Chemistry*, *113*, 1076–1082.

How to cite this article: Henning, S. (2025). Classification of *Kudoa thyrsites* infected and uninfected fish using a handheld near-infrared spectrophotometer, SIMCA and PLS-DA. *Journal of Fish Diseases*, *48*, e14025. https://doi.org/10.1111/jfd.14025