2010-02-01

Fourier Transform Infrared Microspectroscopy and Multivariate Methods for Radiobiological Dosimetry

Aidan Meade
Technological University Dublin, aidan.meade@tudublin.ie

Colin Clarke
Technological University Dublin, Colin.Clarke@tudublin.ie

Hugh Byrne
Technological University Dublin, hugh.byrne@tudublin.ie

See next page for additional authors

Recommended Citation

Meade, A.D. et al (2010) Fourier Transform Infrared Microspectroscopy and Multivariate Methods for Radiobiological Dosimetry. *Radiation Research.* Vol. 173, No. 2, pp. 225-237. doi:10.1667/RR1836.1

Funder: Technology Sector Research (Strand 3) programme of the Irish Higher Education Authority.

Title: Fourier Transform Infrared microspectroscopy and multivariate methods for radiobiological dosimetry

Meade, A.D.[1,2], Clarke[2], C., Byrne, H.J.,[3] and Lyng, F.M.[2]

Running title: FTIRM and multivariate methods for radiobiological dosimetry

1. School of Physics, Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland

2. Radiation and Environmental Science Centre, Focas Institute, Dublin Institute of Technology, Camden Row, Dublin 8, Ireland

3. Focas Institute, Dublin Institute of Technology, Camden Row, Dublin 8, Ireland

ABSTRACT

The scientific literature currently contains an ever growing number of reports of applications of vibrational spectroscopy as a multivariate non-invasive tool for analysis of biological effects at the molecular level. Recently, Fourier Transform Infrared Microspectroscopy (FTIRM) has been demonstrated to be sensitive to molecular events occurring in cells and tissue post-exposure to ionising radiation. In this work the application of FTIRM in the examination of dose dependent molecular effects occurring in skin cells post exposure to ionising radiation, with the use of partial least squares regression (PLSR) and generalized regression neural networks (GRNN) is studied. The methodology is shown to be sensitive to molecular events occurring with radiation dose and time after exposure. The variation in molecular species with dose and time after irradiation is shown to be non-linear by virtue of the higher modelling efficiency yielded from the non-linear algorithms. Dose prediction efficiencies of approximately ±10mGy

have been achieved at 96 hours after irradiation, highlighting the potential applications of the methodology in radiobiological dosimetry.

Introduction

The effect of exposure of biological species to ionising radiation is complex. The physico-chemical reactions that occur along the tracks of energy deposition within the cell generate free radicals and reactive oxygen species, whose subsequent interactions with the membrane, cytoplasm and the nuclear DNA initiate a range of molecular signalling networks that determine the fate of the cell (*1, 2*). Radiation induced DNA lesions such as base deletions, adducts, tandem lesions, single strand and double strand breaks, and clustered damage sites are considered critical to the fate of the cell, and their yield and complexity depends on the dose, dose rate and linear energy transfer of the radiation (*3, 4*). The yield of such damage also depends significantly upon the chemical environment of the DNA, where endogenous thiols (such as cysteamine) and enzymes (such as glutathione, superoxide dismutase) may ameliorate the effect of the radiation exposure through scavenging of free radicals (*3, 5, 6*), while oxygen may assist the formation of DNA lesions through fixation of the damage. In instances where DNA damage is formed, the cell has evolved sophisticated DNA repair mechanisms to deal with the radiation insult (*7, 8*), where the repair mechanism that is adopted by the cell is related to the part of the cell cycle in which it is irradiated (*4*), and may initiate cell cycle arrest (*9*). Effective repair depends on sensing of the damage with specialized proteins and the processing of this damage via specialized repair enzymes (*7, 8*). Deficiencies in

such responses in the cell can lead to increased radiosensitivity (*10*). Despite the complex nature of the network of signalling effects in the cell, some damage is not repaired, possibly due to its level of complexity, which can result in the initiation of cell death via apoptosis (*11*), necrosis, or mitotic cell death (*12*). Other potential outcomes include the senescence of the cell (*13*), autophagic response in which the morphological changes resemble those seen during apoptosis (*14*), in addition to genomic instability and mutagenesis. The level of cell survival after irradiation depends on the level of dose, dose rate, LET of the radation, and on the radiosensitivity of the cell (*4*), with responses such as low dose hypersensitivity and increased radioresistance thought to occur at low doses as a result of variations in the level of sensing of DNA damage (*15*).

The search for non-invasive techniques for biological dosimetry has established the analysis of chromosomal translocations as a signature of radiobiological damage in humans (*16*). Recently, methods which analyse concentration changes in metabolites have also been applied to this problem, identifying biomarkers of radiobiological effect in various biological media (*17-19*). The vibrational microspectroscopy modalities (Raman and FTIR microspectroscopy (FTIRM)) are techniques that can rapidly and non-invasively measure the spatially-resolved chemistry of the cell (*20*) and of tissue (*21*) with minimal sample preparation. FTIRM has proven to be useful in the analysis of complex biological responses during a range of biological processes, including proliferation and cell death processes (*22-25*). Importantly, it has also shown promise as a technique for the examination of the total biochemical content post-irradiation in human cells (*26-28*), human tissue (*29-32*) and bacteria (*33-35*), in addition to isolated molecular

constituents of the cell and its subcellular compartments (*36-38*). Melin and co-workers (*31-33*) have studied the effects of γ-irradiation of D. radiodurans, K. rosea and M. luteus with FTIRM and have demonstrated that differences occurred in spectral signatures associated with nucleic acids, carbohydrates, fatty acids and protein. Gault *et al.* have also demonstrated that changes in the vibrational intensity of bands across the spectrum occur in HaCaT cells after a 6 Gy and 20 Gy γ-radiation dose (*26, 27*), and after a 2Gy dose of alpha particles (*25*).

Infrared radiation, when incident on molecular species, causes transitions within the vibrational manifold of polar bonds i.e. bonds containing an atom with an overall positive charge bonded to one with an overall negative charge. When the frequency of the incident infrared radiation matches that of the natural frequency of vibration of the bond, the bond will absorb the radiation. In complex biological species the absorption bands are broad and represent the superposition of vibrations of various distinct biological species, and may be attributed to families of bond vibrations (eg. lipid or sugar phosphate bond vibrations, carbohydrate osidic bond vibrations etc. (*39, 40*)). A spectrum therefore contains information on the biochemical content of multiple cellular species, and consequently has high dimensionality. Previous studies employing FTIRM in the analysis of chemical changes post-irradiation have simply analysed spectral changes using differences of the mean spectra within each dose class (*31-33*) or analyses of Fourier self-deconvolution spectra (*24-26*), which does not clearly provide information regarding the variation in specific chemical components with dose and time post-irradiation. Such information is provided via the use of multivariate methods (chemometric methods in the

context of modelling spectroscopic data) to analyse the variation in spectral content with exposure and response factors. Many methods have been developed for the analysis of multivariate data (*41*); the choice of any individual method depends on the nature of the multivariate measurement and on the purpose of the analysis, a task which may require some prior knowledge of the system under study. Linear (*42*) and non-linear (*43*) methodologies for regression of multivariate data against various endpoints can provide insights into the underlying chemical effects occurring in the sample with the agent of interest and can produce models which are useful in the prediction of chemical content or effect in new multivariate data (*42, 43*).

In the present study we present for the first time a chemometric analysis of the variation in the content of the FTIR spectra of a human keratinocyte cell line (HaCaT (*44*)) with γ-radiation dose and with time post-irradiation. Since FTIR spectra represent a quantification of the total concentration of organic biochemical species within the cell, this approach ultimately provides radiobiological models which analyse the variation of all the major cellular constituents (lipid, protein, nucleic acid, amino acid, carbohydrate, etc.) simultaneously, which may be used as a means to retrospectively analyse radiation dose. The analysis here employs linear and non-linear modelling techniques that provide the means to predict radiobiological dose. It was found that linear and polynomial multivariate models are not sufficiently comprehensive to describe the variation in the FTIRM biochemical fingerprint with respect to dose and time after irradiation. Features of interest in the predictive performance of these models, and their implications for the application of FTIRM to biological dosimetry, are discussed.

Materials and Methods

Cell Culture

A spontaneously immortalized, aneuploid, cell line was employed in this study, which has been shown to be non-tumourigenic, albeit with a transformed phenotype, in-vitro (*44*). Cells were cultured in Dulbecco's MEM:F12 (1:1) whole medium (Sigma, Dorset, UK) supplemented with 10% fetal calf serum (Gibco, Irvine, UK), 1% penicillin-streptomycin solution 1,000 IU (Gibco, Irvine, UK), 2 mM L-glutamine (Gibco, Irvine, UK) and 1 μg/mL hydrocortisone (Sigma, Dorset, UK). Cells were cultured and maintained in an incubator at 37^{o}C with 95% relative humidity and 5% CO_2. Cells were routinely subcultured at 80%-100% confluency using a 1:1 solution of 0.25% trypsin and 1mM versene at 37^{o}C.

For FTIRM, transmission of the IR radiation through a biological sample is necessary for the acquisition of a spectrum, and therefore spectroscopic substrates that allow the transfer of infrared radiation through the biological sample must be employed. It has previously been demonstrated that FTIRM of HaCaT cells may be performed in transmission-reflection (or 'transreflection') mode using a low emissivity silver oxide coated glass slide (MirrIR, Kevley Technologies), on which a 2% gelatin coating is deposited to enable attachment of the cell and improve cell viability (*22*). Transreflection is essentially a double pass transmission measurement; the broadband IR radiation is

transmitted through the sample to the substrate and is then reflected from the silver oxide coating back through the sample to the collection optics in the FTIR instrument (*22*). In this study MirrIR slides were cut into 20 mm × 25 mm pieces and sterilised in 70% industrial methylated spirits before being placed in 6-well plates and allowed to dry in a laminar flow cabinet. The slides were then washed in phosphate buffer solution (PBS). Approximately 300 μL of a pre-prepared sterile solution of 2% gelatin (b/w) in deionised water (dH_2O) was placed on each substrate, and was stored for 24 hours at 4^oC. After this period, the unbonded portion of the gelatin solution was aspirated from the substrate and the cell suspension was immediately added to the substrate, with the sample subsequently incubated at 37^oC for a further 2-hour period to effect initial attachment. The samples were incubated in 3 ml DMEM:F12 for 24 hours prior to irradiation. For FTIRM analysis at 6 and 12 hours post-irradiation, the cell density used was 1×10^5 cells per substrate, while 5×10^4, 2.5×10^4 and 1.5×10^4 cells per substrate were, respectively, used for analysis at 24 hours, 48 hours and 96 hours post-irradiation.

Irradiation

Three individual passages of the HaCaT cells were used to coat each of three individual substrates for irradiation and analysis at each dose and time point, and were irradiated with γ-rays from a cobalt-60 teletherapy source, which is equipped with a chronometer allowing time settings in 0.01 minute intervals. The dose rate at the sample at the time of irradiation was determined to be 153.47 cGy/min from a decay corrected measurement of the in-beam axial dose at 80 cm source to chamber distance (measured in a water

equivalent phantom using a secondary standard ionization chamber). The dose settings (and the corresponding actual doses delivered with associated uncertainties) were 5 mGy (8.3 ± 40%), 20 mGy (23 ± 13%), 50 mGy (58 ± 17%), 200 mGy (209 ± 5%), 500 mGy (511 ± 2%), 750 mGy (763 ± 1.3%), 1 Gy (1.014 ± 1%), 2.5 Gy (2.514 ± 0.4%) and 5 Gy (5.011 ± 0.2%). The actual dose to which each sample was exposed was determined from the axial dose, corrected for the irradiation time during ingress and egress of the source and the actual irradiation time (a rounding of the calculated irradiation time to 0.01 min increments), scatter and grid factors, together with the source to sample distance. The source to sample distance used was 100 cm with a 30 × 30 cm field size for doses from 5Gy to 50 mGy and 184 cm with a 47 cm × 47 cm field size for the 20 mGy and 5 mGy dose settings. The uncertainty in the dose delivered for each dose point was estimated by assuming an uncertainty in the chronometer setting of ±0.01 min (the minimum chronometer setting).

The dose range for this study was chosen such that a variety of cellular molecular responses would be seen, including those associated with low dose hypersensitivity, increased radioresistance, and apoptotic, necrotic and mitotic cell death. This allows us to then to examine the performance of the models to explain the spectral effects occurring as a result of a range of radiobiological effects. Control samples were sham irradiated and all samples were returned to the incubator immediately after irradiation. After post-irradiation incubation periods of 6, 12, 24, 48 and 96 hours, the cells were fixed in a 4% neutral-buffered formalin solution for 10 minutes, washed three times in dH_2O and dessicated for subsequent FTIRM. Fixation times were kept to a minimum to reduce the

risk of any cellular degradation (*45, 46*). The samples were stored in a dessicator until the time of analysis.

FTIR Microspectroscopy

FTIRM was performed using a Perkin-Elmer GX-II spectrometer. The system is equipped with a mid-infrared source and motorised mid-infrared and far-infrared beamsplitters, allowing the measurement of spectra over the range from 7000 to 50 cm^{-1} with a maximum resolution of 0.3 cm^{-1}. It is also equipped with a ×40 objective, a motorised stage, and a liquid-N_2 cooled MCT detector for operation in either transmission or reflection mode. In the present work spectra were recorded over a 4000 to 720 cm^{-1} wavenumber range with an aperture size of 100 μm ×100 μm, at a spectral resolution of 4 cm^{-1} and with 64 scans per spectrum. All spectra were recorded in transreflection mode; approximately 300 spectra were recorded at each dose and time point.

Pre-processing of FTIR Spectra

Single cell FTIR spectra contain contributions from the chemical content of the sample and physical effects which originate in the optically inhomogeneous nature of the sample under measurement including spurious effects such as scattering within the sample (*47*), resulting in alterations in the baseline. Additional unwanted effects include noise and contamination of the spectra by infrared-active molecules such as water vapour and

carbon-dioxide (*48*). Pre-processing of spectra aims to reduce such contributions, minimising all non-biological variance. Here, all spectral processing and analysis was performed using Matlab 7.2 (The MathWorks Inc., USA) with PLS Toolbox 5.0.3 (Eigenvector Research, Wenatchee, WA, USA).

Firstly outliers within each set of spectra at a given dose and time point were removed using Grubb's test of the Mahalanobis distances between scores of the first three principal components of each spectral set (*49*). Principal components are used to reduce a given multivariate data matrix to a set of orthogonal basis vectors (principal components or eigenvectors) where those vectors with the largest scores (eigenvalues) correspond to the basis spectra which contribute to the largest degree of variance in the dataset (*41, 42*). Subsequently, the contributions of water and carbon dioxide were removed from each spectral set by a modelling procedure using second-derivative spectra of CO_2 and water vapour recorded by the FTIR instrument separately (*48*). Each spectral set was then subjected to the extended multiplicative scatter correction for removal of linear and multiplicative optical effects that contribute to the baseline in the spectra (*47*), and the spectra were then vector normalised to adjust for point-to-point variations in concentration of cells within the aperture window across the sample.

PLS Regression

The partial least squares regression (PLSR) algorithm has found extensive use in the field of chemometrics since its first description by Wold (*50*). The PLSR algorithm constructs

a model that allows the regression of a series of spectral measurements onto target agent or analyte concentrations. The model itself is given as follows (*41*):

$$Y = XB + E \quad (1)$$

where *Y* is a matrix which defines the concentration of agent (in this case radiation dose) associated with each of the multivariate objects (spectra) which themselves are contained within the matrix, *X*, and *E* is a matrix of residuals. The algorithm aims to maximize the association between the structure of X versus the structure of Y and minimize E. An approximation to the procedure by which the algorithm operates visualises the X and Y-matrices being decomposed into their principal components and then regressed against one another. Thus the X and Y-matrices are decomposed into their eigenvectors (principal components, or in the context of PLSR, 'latent variables') and eigenvalues (scores) as:

$$X = T.P^T + E_X \quad (2)$$

$$Y = U.Q^T + E_Y \quad (3)$$

where T and U are the score matrices associated with the transposed matrix of latent variables, P^T and Q^T within the X and Y-matrices respectively, with E_X and E_Y being the residuals associated with the decomposition. The regression model then attempts to approximate the Y-score matrix in equation 3 (*U*) with the X-score matrix, *T*, as:

$$U = TB + H \quad (4)$$

where those latent variables within X, that best regress against Y are retained within the model. Estimates of *U* in equation 5 are then fed into equation 4, providing predictions of Y, $\hat{Y}$, from the X-block data. It is also possible to construct PLSR models that regress the latent variables in X against Y by means of a non-linear model. A nonlinear PLSR model (NL-PLSR) employing a second order polynomial was constructed in an attempt to examine whether the adjustment in the total biochemical content of the cell post-radiation follows a simple non-linear model (which may be associated with cascading or clustering of DNA damage (*51*). The equation describing this model is of the following form (*52*):

$$U = B_0 + B_1T + B_2T^2 + E \quad (5)$$

Generalised Neural Network Regression

Artificial neural networks (ANN's) have been termed 'nonparametric nonlinear regression estimators' (*43*) because of their ability to determine relationships between one or several input or 'independent' variables and one or several output or 'dependent' variables, regardless of the form of the function defining the relationship between the two sets of variables. The network employed here is a generalized regression neural network (GRNN), first described by Specht (*53*), which is a form of kernel regression in which the optimum non-linear regression surface relating the input data to the output data is

determined (*53*). A schematic of the network is presented in figure 1. The network has, as its basis, the following formula which is derived from generalised regression theory (*54*):

$$E\left(y|x\right)=\frac{\int_{-\infty}^{\infty} y.f(x,y).dy}{\int_{-\infty}^{\infty} f(x,y).dy} \qquad (6)$$

where $E\left(y|x\right)$ is the expected value of the output, *y*, for a given input, *x*, and *f(x,y)* is a probability density function of *x* and *y*, which defines the relationship between the two variables. The GRNN effectively estimates *f(x,y)* via a subset of the matrix of the input dataset which is used to train the network. Since the network is developed directly from a training subset of the input dataset, it can generalise to any functional form that relates any input to any target output. The input data is fed through various "layers", which are stages of the network operations and which are arithmetic operations whose parameters are unobservable to the operator. However, key to the operation of the network is the "smoothness" of the radial basis function (RBF) layer (or the width, σ of the RBF distribution functions). The form of these radial basis functions assumes a Gaussian form when σ is large, and may assume non-linear shapes when σ is small. In the first instance, this will increase the prediction error between the output of the network and the desired output but the network will perform well when presented with unseen data, while in the second the prediction error will be small but outliers in the input data will have too great an effect on the regression. Therefore, during training of the network, suitable values for σ (which will be termed the 'smoothness' of the network) are determined through cross-

validation of the training data versus corresponding target outputs; the optimal value for σ is then that value which provides the minimum prediction error (*53, 54*).

Design and Testing of Models

To develop a model of spectral measurements against dose, the spectral dataset was randomized and split into a training set containing 60% of the total number of spectra and a testing set containing the remaining 40% of the spectra. It is known that PLSR models can model noise in the dataset of independent variables if the model chosen is overly complex, i.e. if the number of latent variables chosen is too large. Leave-one-out cross-validation was employed to determine the optimal number of latent variables to retain in the model. The PLSR and NLPLSR algorithms were then executed for ten separate randomisations of the calibration and testing matrices in order to avoid bias in the algorithms when through presentation with different training and testing datasets. In defining the performance of the model in the regression of the spectral measurements against radiation dose at cross-validation we have used the root-mean-squared error of cross-validation (RMSECV). At the model training stage this measure is termed the root-mean-squared error of calibration (RMSEC) and root-mean-squared error of prediction (RMSEP) at the testing stage. The root mean squared error is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(\ddot{Y}_i - Y_i \right)^2}{n}} \qquad (8)$$

where the predicted dose from each model, $\hat{Y}_i$, is compared to the corresponding true value, Y_i. Ten-fold cross-validation was employed to define the optimal smoothness of the GRNN before calibration, in order to prevent over-fitting. The data-set was again randomly split into a training set (60% of total spectra) and unseen testing set (40% of total spectra), and these were used to develop and test the network. Since the network is initialised randomly on each development of the network (i.e. cross-validation, calibration and testing), the network was run over 50 independent initialisations and the mean values of the RMSEC and RMSEP were computed to evaluate the performance of the GRNN. Execution of the models repeatedly in this manner avoids data bias by randomisation of the training and testing data matrices. The values of RMSEP quoted in this work are conservative estimates due to the uncertainties in the dose delivered (estimates of which are included in the section on *Irradiation* earlier), a factor that has been demonstrated to lower the actual RMSEP (*55, 56*).

PLSR, and NL-PLSR models were evaluated separately for corrected spectra, 1st order derivative spectra and second order derivative spectra. First order derivatives are employed in spectroscopy for the removal of slowly varying background or baseline features below chemical signatures, and are employed here in a separate modelling procedure as a check of the preprocessing procedures that were employed to correct the spectra. Second order derivative spectra are normally employed to highlight 'shoulder' features occurring on spectral peaks, and are here employed to determine whether such features are important in terms of their relationship to dose or in the ability to predict radiation dose from such spectra.

Results

Spectral Vibrations in HaCaT cells

The main FTIRM spectral features of sham-irradiated and irradiated HaCaT cells at 96 hours post-irradiation are shown in figure 2. The associated band assignments are shown in table 1 (with associated references to their sources). Mean spectra from each spectral category are presented here to highlight spectral changes occurring as a result of exposure to ionising radiation, although a full interpretation of the multivariate models involves detailed analysis of their fitting parameters. Detailed analyses of the correlation of the complex changes to the spectral response as a function of dose and exposure time with gold standard tests for viability, proliferative capacity and mitochondrial activity are currently underway.

In sham-irradiated cells, the region up until 1750 cm^{-1} (fingerprint region) shows vibrations of characteristic modes of DNA and RNA (phosphate backbone, C-O stretch, C-N-C stretch, amide and characteristic modes of certain nucleic acid bases), lipid (-CH_3 and -CH_2 bending and scissoring and COO stretch), and protein (carbonyl stretch coupled to N-H bending vibration of peptide backbone (amide I), C-N stretch coupled to N-H bend (amide II) and C=O bend coupled to N-H bending vibration (amide III)). In the higher wavenumber region (from ~ 2600-3500 cm^{-1}), the vibrational modes are dominated by methyl -CH_2 and -CH_3 stretching vibrations in lipid and protein, together

with interaction of the excited state of the N-H stretching vibration and the overtone of the amide II vibration (amide A and amide B vibration respectively). Stretching vibrations of the O-H bond occur in both the fingerprint and higher wavenumber region, and have origin in water contamination of the sample, or O-H bond vibrations within carbohydrates (*57, 58*).

Cross Validation, Calibration and Prediction Performance of PLSR, NL-PLSR and GRNN Models

The complexity of a partial least squares (PLSR) model is determined by the number of latent variables incorporated in the model (which are orthogonal dimensions that explain the variation in the multivariate dataset with respect to the target variable). Selection of a suitable number of latent variables to retain in constructing the PLSR and NL-PLSR models is essential to ensure that over-fitting of the data does not occur (*42*). Martens and Naes have defined a condition indicating the optimal number of latent variables to retain in such models as the minimum of a plot of the cross-validation error versus the latent variable number in an independent test set (*42*). Adopting this approach we have determined that the optimal number of latent variables (nLVs) for both the PLSR and NL-PLSR models is as shown in tables 2 and 3, with examples of the corresponding cross-validation, and subsequent calibration and testing, results shown in figure 4(a) and (b).

Tables 2 and 3 also show the corresponding variance explained by the first two latent variables for the PLSR and NL-PLSR models at each time point, and the corresponding root mean squared errors of calibration (RMSEC) and prediction (RMSEP). The explained variance with respect to LV above the second LV is not shown as these describe decreasing amounts of the variance in the dataset with respect to the target dose, and the variance described by the first two LV's is quite high (above 96-97% for both the corrected and 1st order derivative spectra) in each model. There is good agreement between the values of RMSEC and RMSEP for each PLSR and NL-PLSR model at each dose point, indicating that overfitting is not evident in the modelling results, despite the large number of LV's required to model the data. In addition, the values of RMSEC and RMSEP for the models developed with either corrected spectra or 1st order derivative spectra agree well with one another, indicating that the spectral pre-processing employed here is effective at removing spectral variance within baseline features, and that such effects have little influence on the predictive efficiency of the models.

Table 4 shows the performance of the GRNN in training (calibration) and prediction of dose, and the associated goodness-of-fit statistics. An example of the results seen at cross-validation of the network, and subsequent calibration and testing, are shown in figure 4(c). Also included in the table are the values of the network smoothing factor. It is notable that the value of the smoothing factor of the neural network is quite similar for the network developed for each spectral dataset at each time point. This indicates that the range of spectral features that are modelled with each independent neural network are quite consistent, even though the weights of each network used to combine the spectral

features in regressing against dose may be different at each training stage. Again the values of RMSEC and RMSEP for each of the networks are quite similar, indicating that the networks do not over-fit spectral features within the datasets, and as such generalise well to previously unseen data.

Discussion

Spectral Features Varying with Dose

There is a paucity of studies on the use of vibrational spectroscopy for the analysis of biological samples exposed to ionising radiation. Although studies with FTIRM by Melin and co-workers (*31-33*) on the effects of γ-irradiation of D. radiodurans, K. rosea and M. luteus are not directly comparable to effects on eukaryotic cells, they demonstrate that spectral differences occur in the 1245 to 900 cm^{-1} (nucleic acids and carbohydrates), 3100-2800 cm^{-1} (fatty acids and protein) and 1750 to 1390 cm^{-1} (lipid and protein) regions of the spectrum (in certain bacterial strains these band intensities were observed to increase, while in others they were observed to decrease). Gault *et al.* have also demonstrated that changes in vibration intensity of bands across the fingerprint region occur in HaCaT cells after a 6 Gy and 20 Gy γ-radiation dose, and after a 2Gy dose of alpha particles (*25, 65*).

A straightforward means of examining the spectral changes occurring after irradiation is through examination of the difference spectra between the mean sham-irradiated cell

spectra and those for the irradiated cells. An example of the difference spectra seen for 500mGy and 5Gy dose categories at 96 hours post irradiation is shown in Figures 3(a) and 3(b). It must be noted in interpreting these difference spectra that a peak denotes an absorbance in the sham-irradiated cell spectra that is higher than the corresponding absorbance in the irradiated sample, and vice versa. The findings of Gault et al are supported by the features in figure 3(a) and 3(b) of the present work. Absorbance differences between 1200 cm^{-1} and 1030 cm^{-1} are due to vibrations of osidic (C-O) bond stretching in DNA, RNA and other carbohydrates, together with overlapping O-H deformation vibrations in carbohydrates from 1290 to 1030 cm^{-1} (*57-59*). There are also differences attributable to changes in asymmetric (~1230 cm^{-1}) and symmetric (~1090-1084 cm^{-1}) stretching of phosphate bond vibrations in RNA and DNA. Other characteristic vibrations in nucleic acids (bending vibration in uracil at 996 cm^{-1} and stretching vibration in PO_4^- moieties in DNA) also exhibit slight absorbance differences. In addition there is a strong absorbance difference in the tyrosine ring vibration band at ~1515 cm^{-1} in the 500mGy sample. As a complete picture these features could signify an increase in DNA strand breaks and base cleavage reactions with dose, increasing hydroxylation of C=C double bonds in purine and pyrimidine rings of DNA, and potentially, the formation of DNA-DNA or DNA-protein cross-links (*25, 26*). The absorbance differences at 1160 cm^{-1} (-C-OH in nucleic acid carbohydrates) and 1242 cm^{-1} (-PO_2^- asymmetric stretching vibration) also suggest adjustments to the hydrogen bonding structure in DNA that have been seen previously in apoptosis in irradiated lymphocytes (*26*). In addition, the pyranose carbohydrate vibrations (symmetric and asymmetric ring vibrations, C-H deformation) exhibit positive absorbance differences

from 935 to 905 cm^{-1}, while the O-H bond stretching vibration has highly positive absorbance differences at ~3176 cm^{-1} and 3262 cm^{-1}. These features could signify an adjustment to metabolic activity involving glucose synthesis for energy production in molecular transduction responses to radiation exposure (*15, 60*).

The absorbance differences in protein amide bands suggest the occurrence of dose dependent secondary structural changes to protein, previously seen by Gault et al (*27, 28*). Such changes to the structure of protein have been demonstrated to result from chain cleavage, formation of protein-protein cross-links and amino acid degradation post-ionising radiation exposure (*25, 26*). However, many of the regions of the spectrum ascribed to protein also contain vibrations that are assigned to lipid species, to which the highlighted signatures may be assigned. This could be evidence of degradation to the structure and function of biomembranes within the cell (*25, 26*). There are also absorbance differences in the CO-O-C symmetric and symmetric stretching vibration at 1170 and 1070 cm^{-1} in lipids, which supports this latter assertion.

Comparative Predictive Efficiency of Modelling Procedures

The values of RMSEC and RMSEP for the PLSR, NL-PLSR and GRNN models may be compared through consulting the top panels of table 2 and 3 with table 4. It is clear that both the PLSR and NL-PLSR models have similar performances in prediction of dose from spectral information, and generalise well to unseen data, although the NL-PLSR model is slightly less effective in comparison with the PLSR model. The GRNN,

however, outperforms both the PLSR and NL-PLSR models in terms of prediction and generalisation ability, achieving values of RMSEC and RMSEP that are in general lower than the other two models. It is also notable that for the PLSR and NL-PLSR models, the use of 2nd order derivative spectra shows no improvement in the predictive ability of each model, but rather appears to disimprove them.

The large difference between the RMSEC and RMSEP values obtained using the GRNN versus either the PLSR or NL-PLSR models suggest that the concentration of many chemical components of the cell vary in a non-linear manner with dose, and this variation is not explained adequately via either a linear or second order polynomial model. This demonstrates that subtle and complex dose-dependent changes are apparent in FTIR spectral fingerprints.

The inter-comparison of RMSEC and RMSEP values between GRNN models at different time points can give insights into the dynamics of the system under study (see table 4 and 5). The predictive efficiency of the PLSR and NL-PLSR models are relatively consistent with time post-irradiation, while the predictive efficiency of the GRNN varies with time post-irradiation, being best at 6 hours and 96 hours post-irradiation, falling from a maximum at 12 hours post-irradiation onwards. This suggests that there is an increased variance in the biochemical content of the cell population at 12 hours post-irradiation, suggesting that the total chemical composition of the cells is varying as a result of the initiation of a variety of DNA repair, and other, response mechanisms (*15, 60*). This response appears to induce a range of spectral content profiles across the population of

cells while they undergo responses such as DNA repair, apoptosis or necrosis (*15, 60*). It is also possible that this increase in variance within the population of cells is a result of the difference in response mechanisms occurring within cells exposed to doses that may initiate non-targeted effects (from 5 mGy to 200 mGy (*15*)) in comparison to those exposed to doses producing effects consistent with the classical paradigm (*3, 15*). The reduction in the RMSEC and RMSEP values at 24 hours suggests that spectral content and thus dose-dependent modelling efficiency may also be affected by the synchronisation of cells within the cell cycle (*9*) (the length of the cell cycle in HaCaT cells is approximately 23 hours (*44*)). At 96 hours, it appears that cells within a population have exhausted the range of molecular response mechanisms available to them, since variance in the spectral content within each dose category reduces, and the predictive efficiency of the modelling procedures increases; this suggests that the spectral content within the population of cells varies according to a more consistent model with respect to radiation dose. The predictive efficiency of the model at this time point is excellent (approximately ± 10 mGy including the variance in RMSEP) when compared to that achieved with established methods of biological dosimetry. This RMSEP equates to a measurement uncertainty which varies between 200% for a dose of 5 mGy, to 2% at 500 mGy and 0.2% at 5Gy, while this uncertainty has been quoted as high as 62% for estimates based on measurements of chromosomal translocation frequency (*61*).

Finally, the PLSR and GRNN algorithms were applied to the development of models based on the complete dataset of spectra at all time points. The PLSR algorithm in this case is termed a PLS2 algorithm (*42*), and predicts both the dose and time of irradiation.

The RMSEC and RMSEP for prediction of time from this algorithm were both 17.9 hours (with R^2 values of approximately 0.85 and 0.84 respectively), while the corresponding values for RMSEC and RMSEP in prediction of dose were both 1.3 Gy (with R^2 values of 0.41 and 0.43 respectively). In contrast, the GRNN was capable of predicting time to within 6.3 hours (with an RMSEC of 5.8 hours and R^2 values of 0.99 and 0.98 at calibration and testing respectively) and dose to within 0.37 Gy (with an RMSEC of 0.25 and R^2 values of 0.99 and 0.97 at calibration and prediction respectively). This emphasizes the degree to which chemical signatures vary in a non-linear manner with dose and time post irradiation.

Conclusion

This study demonstrates that FTIRM, in addition to its potential in cytometry and tissue pathology, provides a platform form the non-invasive measurement of radiobiological damage as it is sensitive to the complex series of molecular responses produced in the cell. It has been demonstrated that powerful multivariate techniques can offer the means to analyse the changes in the biochemical fingerprint occurring with dose and time after irradiation as a platform for retrospective biological dosimetry. The study raises questions regarding the nature of the non-linearities in these changes that are suggested by the performance of the GRNN in modelling the biochemical fingerprint. Further detailed investigations are currently being directed towards interpreting the modelling parameters within the multivariate models, with a view to elucidating the identity of radiobiologically relevant spectral features and modelling their association with

radiobiological effect. A correlation of the observed spectral responses as a function of radiation dose with biochemical endpoints could potentially add considerable insight into the molecular origin of the response.

Acknowledgements

The authors acknowledge funding through the Technology Sector Research (Strand 3) programme of the Irish Higher Education Authority. The Focas Institute, DIT has been established under the Irish HEA Programme for Research in Third Level Institutions, Cycle 1 (1999-2001). The authors would like to thank Professor Harald Martens of the University of Copenhagen, Denmark for his kind advice. The authors also thank St. Luke's Hospital Dublin for continued support.

References

1. R. B. Mikkelsen and P. Wardman, Biological chemistry of reactive oxygen and nitrogen and radiation-induced signal transduction mechanisms. *Oncogene* **22**, 5734-5754 (2003).
2. P. Dent, A. Yacoub, J. Contessa, R. Caron, G. Amorino, K. Valerie, M. P. Hagan, S. Grant and R. Schmidt-Ullrich, Stress and radiation-induced activation of multiple intracellular signaling pathways. *Radiation research* **159**, 283-300 (2003).
3. E. J. Hall and A. J. Giaccia, *Radiobiology for the Radiologist*. Lippincott Williams and Wilkins, 2005.
4. B. G. Wouters and A. C. Begg, Irradiation-induced damage and the DNA damage response. In *Basic Clinical Radiobiology* (M. C. Joiner and A. van der Kogel, Eds.), pp. 11-26. Hodder Arnold, London, 2009.
5. P. O'Neill, Radiation-induced damage in DNA. In *Radiation Chemistry - Present status and future trends* (C. D. Jonah and B. S. Madhava Rao, Eds.), pp. 585-622. Elsevier, London, 2001.
6. C. von Sonntag, *The Chemical Basis of Radiation Biology*. Taylor and Francis, London, 1987.

7. L. Brugmans, R. Kanaar and J. Essers, Analysis of DNA double-strand break repair pathways in mice. *Mutation research* **614**, 95-108 (2007).
8. J. Falck, J. Coates and S. P. Jackson, Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**, 605-611 (2005).
9. M. B. Kastan and J. Bartek, Cell-cycle checkpoints and cancer. *Nature* **432**, 316-323 (2004).
10. I. Szumiel, Intrinsic radiation sensitivity: cellular signaling is the key. *Radiation research* **169**, 249-258 (2008).
11. R. C. Taylor, S. P. Cullen and S. J. Martin, Apoptosis: controlled demolition at the cellular level. *Nature reviews* **9**, 231-241 (2008).
12. M. Castedo, J. L. Perfettini, T. Roumier, K. Andreau, R. Medema and G. Kroemer, Cell death by mitotic catastrophe: a molecular definition. *Oncogene* **23**, 2825-2837 (2004).
13. J. Campisi and F. d'Adda di Fagagna, Cellular senescence: when bad things happen to good cells. *Nature reviews* **8**, 729-740 (2007).
14. D. J. Klionsky, Autophagy: from phenomenology to molecular understanding in less than a decade. *Nature reviews* **8**, 931-937 (2007).
15. K. M. Prise, G. Schettino, M. Folkard and K. D. Held, New insights on cell death from radiation exposure. *The lancet oncology* **6**, 520-528 (2005).
16. J. D. Tucker, Low-dose ionizing radiation and chromosome translocations: a review of the major considerations for human biological dosimetry. *Mutation research* **659**, 211-220 (2008).
17. A. D. Patterson, H. Li, G. S. Eichler, K. W. Krausz, J. N. Weinstein, A. J. Fornace, Jr., F. J. Gonzalez and J. R. Idle, UPLC-ESI-TOFMS-based metabolomics and gene expression dynamics inspector self-organizing metabolomic maps as tools for understanding the cellular response to ionizing radiation. *Analytical chemistry* **80**, 665-674 (2008).
18. G. Rainaldi, R. Romano, P. Indovina, A. Ferrante, A. Motta, P. L. Indovina and M. T. Santini, Metabolomics using 1H-NMR of apoptosis and Necrosis in HL60 leukemia cells: differences between the two types of cell death and independence from the stimulus of apoptosis used. *Radiation research* **169**, 170-180 (2008).
19. J. B. Tyburski, A. D. Patterson, K. W. Krausz, J. Slavik, A. J. Fornace, Jr., F. J. Gonzalez and J. R. Idle, Radiation metabolomics. 1. Identification of minimally invasive urine biomarkers for gamma-radiation exposure in mice. *Radiation research* **170**, 1-14 (2008).
20. C. Matthaus, S. Boydston-White, M. Miljkovic, M. Romeo and M. Diem, Raman and infrared microspectral imaging of mitotic cells. *Appl Spectrosc* **60**, 1-8 (2006).
21. F. M. Lyng, E. O. Faolain, J. Conroy, A. D. Meade, P. Knief, B. Duffy, M. B. Hunter, J. M. Byrne, P. Kelehan and H. J. Byrne, Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool. *Exp Mol Pathol* **82**, 121-129 (2007).
22. A. D. Meade, F. M. Lyng, P. Knief and H. J. Byrne, Growth substrate induced functional changes elucidated by FTIR and Raman spectroscopy in in-vitro cultured human keratinocytes. *Anal Bioanal Chem* **387**, 1717-1728 (2007).

23. M. Diem, M. Romeo, S. Boydston-White, M. Miljkovic and C. Matthaus, A decade of vibrational micro-spectroscopy of human cells and tissue (1994-2004). *The Analyst* **129**, 880-885 (2004).
24. K. Z. Liu, L. Jia, S. M. Kelsey, A. C. Newland and H. H. Mantsch, Quantitative determination of apoptosis on leukemia cells by infrared spectroscopy. *Apoptosis* **6**, 269-278 (2001).
25. J. Mourant, P. Kunapareddy, S. Carpenter and J. P. Freyer, Vibrational spectroscopy for identification of biochemical changes accompanying carcinogenesis and the formation of necrosis. *Gynecol Oncol* **99**, S58-60 (2005).
26. N. Gault and J. L. Lefaix, Infrared microspectroscopic characteristics of radiation-induced apoptosis in human lymphocytes. *Radiation research* **160**, 238-250 (2003).
27. N. Gault, O. Rigaud, J. L. Poncy and J. L. Lefaix, Infrared microspectroscopy study of gamma-irradiated and H2O2-treated human cells. *Int J Radiat Biol* **81**, 767-779 (2005).
28. N. Gault, O. Rigaud, J. L. Poncy and J. L. Lefaix, Biochemical alterations in human cells irradiated with alpha particles delivered by macro- or microbeams. *Radiation research* **167**, 551-562 (2007).
29. R. Jyothi Lakshmi, V. B. Kartha, C. Murali Krishna, R. S. JG, G. Ullas and P. Uma Devi, Tissue Raman spectroscopy for the study of radiation damage: brain irradiation of mice. *Radiation research* **157**, 175-182 (2002).
30. A. Synytsya, P. Alexa, J. Besserer, J. De Boer, S. Froschauer, R. Gerlach, M. Loewe, M. Moosburger, I. Obstova, et al., Raman spectroscopy of tissue samples irradiated by protons. *Int J Radiat Biol* **80**, 581-591 (2004).
31. M. S. Vidyasagar, K. Maheedhar, B. M. Vadhiraja, D. J. Fernendes, V. B. Kartha and C. M. Krishna, Prediction of radiotherapy response in cervix cancer by Raman spectroscopy: a pilot study. *Biopolymers* **89**, 530-537 (2008).
32. J. de Boer, A. Synytsya, P. Alex, J. Besserer, S. Froschauer, M. Loewe, M. Moosburger, K. Volka and M. Wurkner, The effect of proton-irradiation on the Raman spectroscopy of tissue samples. *Radiother Oncol* **73 Suppl 2**, S102-104 (2004).
33. A. M. Melin, A. Perromat and G. Deleris, Sensitivity of Deinococcus radiodurans to gamma-irradiation: a novel approach by Fourier transform infrared spectroscopy. *Archives of biochemistry and biophysics* **394**, 265-274 (2001).
34. A. M. Melin, A. Perromat, C. Lorin and G. Deleris, Gamma irradiation and cellular damage in Kocuria rosea: investigation by one- and two-dimensional infrared spectroscopy. *Archives of biochemistry and biophysics* **408**, 211-219 (2002).
35. A. Perromat, A. M. Melin, C. Lorin and G. Deleris, Fourier transform IR spectroscopic appraisal of radiation damage in Micrococcus luteus. *Biopolymers* **72**, 207-216 (2003).
36. K. Sailer, S. Viaggi and M. Nusse, Radiation-induced structural modifications in dsDNA analysed by FT-Raman spectroscopy. *Int J Radiat Biol* **69**, 601-613 (1996).
37. K. Sailer, S. Viaggi and M. Nusse, Kinetics of radiation- and cytochrome c-induced modifications in liposomes analysed by FT-Raman spectroscopy. *Biochim Biophys Acta* **1329**, 259-268 (1997).
38. S. P. Verma, Low levels of irradiation modify lipid domains in model membranes: a laser Raman study. *Radiation research* **107**, 183-193 (1986).

39. P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectroscopy*. John Wiley and Sons, Hoboken, New Jersey, 2007.
40. D. Naumann, FT-Infrared and FT-Raman Spectroscopy in Biomedical Research. In *Infrared and Raman Spectroscopy of Biological Materials* (H. U. Gremlich and B. Yan, Eds.), pp. 323-377. Marcel Dekker Inc., New York, 2001.
41. K. Varmuza and P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press, Taylor and Francis Group, Boca Raton, FL, 2009.
42. H. Martens and T. Naes, *Multivariate Calibration*. John Wiley and Sons, 1994.
43. F. Despagne and D. L. Massart, Neural networks in multivariate calibration. *The Analyst* **123**, 157R-178R (1998).
44. P. Boukamp, R. T. Petrussevska, D. Breitkreutz, J. Hornung, A. Markham and N. E. Fusenig, Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J Cell Biol* **106**, 761-771 (1988).
45. E. Gazi, J. Dwyer, N. P. Lockyer, J. Miyan, P. Gardner, C. Hart, M. Brown and N. W. Clarke, Fixation protocols for subcellular imaging by synchrotron-based Fourier transform infrared microspectroscopy. *Biopolymers* **77**, 18-30 (2005).
46. G. Hastings, R. Wang, P. Krug, D. Katz and J. Hilliard, Infrared microscopy for the study of biological cell monolayers. I. Spectral effects of acetone and formalin fixation. *Biopolymers* **89**, 921-930 (2008).
47. A. Kohler, C. Kirschner, A. Oust and H. Martens, Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin. *Appl Spectrosc* **59**, 707-716 (2005).
48. S. W. Bruun, A. Kohler, I. Adt, G. D. Sockalingum, M. Manfait and H. Martens, Correcting attenuated total reflection-Fourier transform infrared spectra for water vapor and carbon dioxide. *Appl Spectrosc* **60**, 1029-1039 (2006).
49. F. Grubbs, Procedures for detecting outlier observations in samples. *Technometrics* **11**, 1-21 (1969).
50. S. Wold, H. Martens and H. Wold, The multivariate calibration problem in chemistry solved by the PLS method. In *Proc. Conf. Matrix Pencils*, pp. 286-293. Springer Verlag, Heidelberg, 1983.
51. P. O'Neill and P. Wardman, Radiation chemistry comes before radiation biology. *International Journal of Radiation Biology* **85**, 9-25 (2009).
52. F. Dieterle, S. Busche and G. Gauglitz, Different approaches to multivariate calibration of nonlinear sensor data. *Anal. Bioanal. Chem.* **380**, 383-396 (2004).
53. D. F. Specht, A general regression neural network. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* **2**, 568-576 (1991).
54. P. D. Wasserman, *Advanced Methods in Neural Computing*. Van Nostrand Reinhold, New York, 1993.
55. A. C. Olivieri, N. K. M. Faber, J. Ferre, R. Boque, J. H. Kalivas and H. Mark, Uncertainty estimation and figures of merit for multivariate calibration. *Pure Appl. Chem.* **78**, 633-661 (2006).
56. J. A. F. Pierna, L. Jin, F. Wahl, N. M. Faber and D. L. Massart, Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error. *Chemometrics and Intelligent Laboratory Systems* **65**, 281-291 (2003).

57. G. Socrates, *Infrared and Raman Characteristic Group Frequencies - Tables and Charts*. John Wiley and Sons, Chichester, England, 2001.
58. D. Naumann, FT-Infrared and FT-Raman Spectroscopy in Biomedical Research. In *Infrared and Raman Spectroscopy of Biological Materials* (H. U. Gremlich and B. Yan, Eds.), pp. 323-378, 2001.
59. E. Taillandier and J. Liquier, Infrared spectroscopy of DNA. *Methods in enzymology* **211**, 307-335 (1992).
60. B. F. Trump and I. K. Berezesky, The Reactions of Cells to Injury: Oncosis and Necrosis - The Role of Calcium. In *When Cells Die. A Comprehensive Evaluation of Apoptosis and Programmed Cell Death* (R. A. Lockshin, Z. Zakeri and J. L. Tilly, Eds.), pp. 57-96. Wiley-Liss, New York, 1998.
61. P. Bhatti, D. L. Preston, M. M. Doody, M. Hauptmann, D. Kampa, B. H. Alexander, D. Petibone, S. L. Simon, R. M. Weinstock, et al., Retrospective biodosimetry among United States radiologic technologists. *Radiation research* **167**, 727-734 (2007).
62. M. Diem, Romeo, M, Matthäus, C, Miljkovica, M, Miller, L, Lasch, L, Comparison of Fourier transform infrared (FTIR) spectra of individual cells acquired using synchrotron and conventional sources. *Infrared Physics and Technology* **45**, 331-338 (2004).
63. H. Y. N. Holman, Martin, M.C., Blakely, E.A., Bjornstad, K., McKinney, W.R., IR spectroscopic characteristics of cell cycle and cell death probed by synchrotron radiation based fourier transform IR spectromicroscopy. *Biopolymers (Biospectroscopy)* **57**, 329-335 (2000).
64. P. Lasch, A. Pacifico and M. Diem, Spatially resolved IR microspectroscopy of single cells. *Biopolymers* **67**, 335-338 (2002).
65. A. Tfayli, O. Piot, A. Durlach, P. Bernard and M. Manfait, Discriminating nevus and melanoma on paraffin-embedded skin biopsies using FTIR microspectroscopy. *Biochim Biophys Acta* **1724**, 262-269 (2005).

Table Captions:

Table 1. Tentative FTIR vibrational assignments (for cellular spectra) taken from various sources ((*26, 27*), (*40, 62-65*) and references therein). ν = bond stretch; s = symmetric vibration; as = asymmetric vibration; δ = bending vibration; sc = scissoring vibration; p = protein; l = lipid;

Table 2. Performance of PLSR model in prediction of dose at each time point post-irradiation ((RMSEC and RMSEP are in the units of dose (Gy)). All values presented are the mean of ten independent executions of the model. 'nLv' denotes the number of latent variables required to explain the spectral variance (see the text of the manuscript for a description of this). The variance described indicates the percentage of the variance in the dataset explained by the model (in a specific Latent variable (Lv)). Analyses have been performed separately for raw pre-processed (i.e. corrected) spectra and their first and second order derivatives. Figures in brackets denote the standard deviations on the mean

Table 3. Performance of NL-PLSR model in prediction of dose at each time point post-irradiation ((RMSEC and RMSEP are in the units of dose (Gy)). All values presented are the mean of ten independent executions of the model. 'nLv' denotes the number of latent variables required to explain the spectral variance (see the text of the manuscript for a description of this). The variance described indicates the percentage of the variance in the dataset explained by the model (in a specific Latent variable (Lv)). Analyses have been performed separately for raw pre-processed (i.e. corrected) spectra and their first and second order derivatives. Figures in brackets denote the standard deviations on the mean

Table 4. Performance of GRNN in regression of spectral measurements against radiation dose ((RMSEC and RMSEP are in the units of dose (Gy)) at each time point. Each value of RMSEC or RMSEP is the mean of 50 independent executions of the network for complete randomizations of the training and testing datasets. GRNN's were generated on the raw spectral data at each time point post irradiation. The 'smoothing' factor is the value of 'n' required in the radial basis function (RBF) of the neural network (see the text of the manuscript for a description of this). Figures in brackets denote the standard deviations on the mean

Figure Captions:

Figure 1. Schematic representation of neural network architecture used for regression of spectral measurements against radiation dose.

Figure 2. Mean FTIR spectra of HaCaT cells exposed to 0.5Gy, 5Gy, and sham-irradiated, at 96 hours post-irradiation. The width of the trace in each case denotes the extent of the standard errors for each spectral category.

Figure 3(a). Difference spectra in fingerprint region between mean sham-irradiated and mean irradiated HaCaT cells in two dose categories at 96 hours post-irradiation. Absorbance changes seen in regions assigned to carbohydrate, nucleic acid, protein and lipid. Standard errors are shown as shaded bands around the difference spectra.

Figure 3(b). Difference spectra in high wavenumber region between mean sham-irradiated and mean irradiated HaCaT cells in two dose categories at 96 hours post-irradiation. Absorbance changes seen in regions assigned to carbohydrate, protein and lipid. Standard errors are shown as shaded bands around the difference spectra.

Figure 4. Prediction and cross-validation performance of (a) PLS, (b) NL-PLS and (c) GRNN models at 96 hours post-irradiation. The top panel depicts the cross-validation performance of each of the models in selection of the optimal model complexity. The bottom panel displays the test performance of each of the models in estimating the radiation dose from spectral measurements versus the actual dose delivered. The reader should note that the uncertainty in 'actual' dose delivered is within the bounds defined by the dimensions of the symbols on the graph.

Tables

Table 1:

Wavenumber (cm^{-1})	Assignment
3520-3100	ν O-H (carbohydrate)
~3290	Amide A (ν -N-H), (p)
3200-3000	ν as $–NH_3^+$ (free amino acids)
3000-2850	ν C-H (free amino acids)
~3100	Overtone of Amide II band
~3050	Amide B (ν -N-H), p
3030-3020	ν as $-CH_3$ (l)
~3010	ν =C-H (l)
~2960	ν as $-CH_3$ (l, p)
~2920	ν as $-CH_3$ (l, p)
~2875	ν s $-CH_3$ (l, p)
~2850	ν s $–CH_2$ (l, p)
1720-1745	ν –C=O (l) (esters)
1710-1716	ν as –C=O (RNA, esters)
1705-1690	ν as –C=O (RNA, DNA)
1654	Amide I ν -C=O (80%), ν - C-N (10%), δ -N-H (10%), α-helix
1640-1630	Amide I ν -C=O (80%), ν - C-N (10%), δ -N-H (10%), β-sheet
1610, 1578	ν -C4-C5, ν -C=N (imidazole ring, DNA, RNA)
1550-1540	Amide II δ -N-H (60%), ν - C-N (40%), α-helix
1530	Amide II δ -N-H (60%), ν - C-N (40%), β-structure
1515	Aromatic tyrosine ring
1467	δ $–CH_2$ (l, p)
1455	δ as $–CH_3$ / $-CH_2$, sc (l, p)
1400-1370	ν $-COO^-$, δ s $–CH_3$ (l, p)
1330-1200	Amide III (p)
1290-1030	O-H def., (carbohydrate)
1244-1230	ν as $–PO_2^-$ (RNA, DNA)
1200-1030	ν -C-O (pyranose carbohydrates)
1160-1000	ν -C-O (carbohydrate)
1160, 1120	ν -C-O (RNA ribose)
1170, 1070	ν as, ν s –CO-O-C (l)
1090-1084	ν s $–PO_2^-$ (RNA, DNA)
1060, 1050	ν –C-O (deoxyribose/ribose DNA, RNA)
996	ν s (RNA and δ ring of uracyl)
965	ν s - PO_4^- (DNA and deoxyribose-phosphate skeletal motions)
975-960	Terminal methyl def. (amino acids)
960-730	C-H deformation (pyranose carbohydrates)
935-905	Asymmetric ring vibration (pyranose carbohydrates)
900-800	ν s CNC (p)
855-835	C-H def. (α-pyranose carbohydrates)
785-760	Ring vibration (β-pyranose carbohydrates)

Table 2:

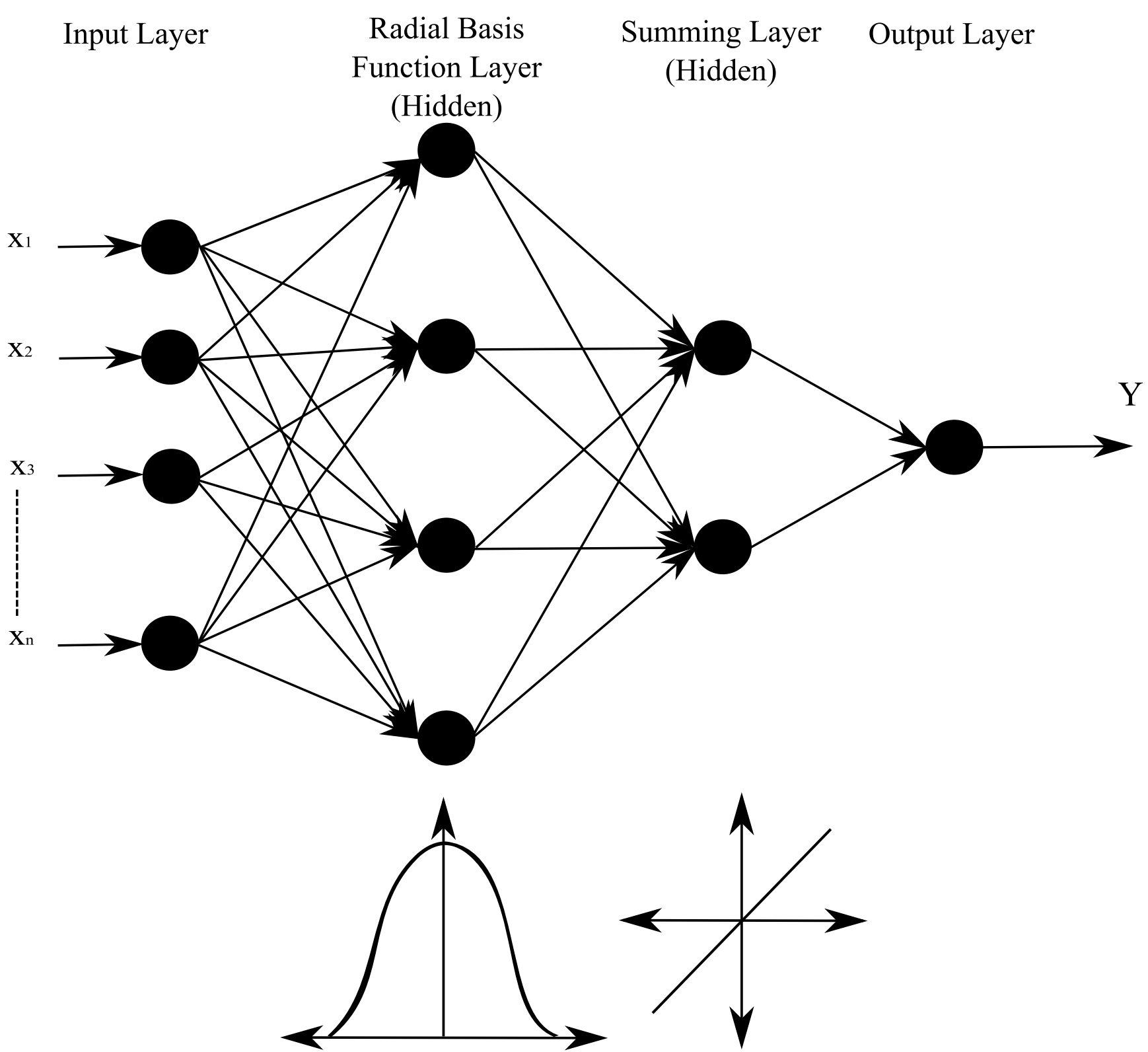
Time	nLv	RMSEC	RMSEP	R^2 Cal.	R^2 Test	Variance Described (LV 1)	Variance Described (LV 2)
Corrected Spectra							
6	22	0.22 (0.01)	0.31 (0.02)	0.99	0.98	96.6	2.5
12	21	0.59 (0.03)	0.79 (0.03)	0.92	0.87	97.2	2.1
24	21	0.24 (0.02)	0.33 (0.02)	0.99	0.98	97.4	2.3
48	23	0.32 (0.01)	0.46 (0.02)	0.97	0.95	95.4	3.6
96	20	0.27 (0.02)	0.37 (0.01)	0.98	0.96	94.7	1.6
1st Order Derivative							
6	13	0.25 (0.03)	0.37 (0.01)	0.98	0.97	94.3	2.5
12	16	0.43 (0.09)	0.89 (0.07)	0.96	0.83	96.1	1.9
24	18	0.18 (0.05)	0.51 (0.02)	0.99	0.94	96.9	2.05
48	20	0.19 (0.1)	0.59 (0.03)	0.99	0.91	95.3	2.9
96	15	0.22 (0.02)	0.44 (0.01)	0.99	0.94	95.7	2.3
2nd Order Derivative							
6	11	0.32 (0.04)	0.51 (0.03)	0.97	0.93	72.7	16.3
12	11	0.63 (0.15)	1.14 (0.04)	0.91	0.67	70.1	13.9
24	11	0.52 (0.12)	0.90 (0.03)	0.94	0.81	80.7	9.8
48	11	0.52 (0.06)	0.92 (0.05)	0.93	0.78	88.5	4.6
96	10	0.49 (0.12)	0.88 (0.04)	0.92	0.76	64.5	18.3

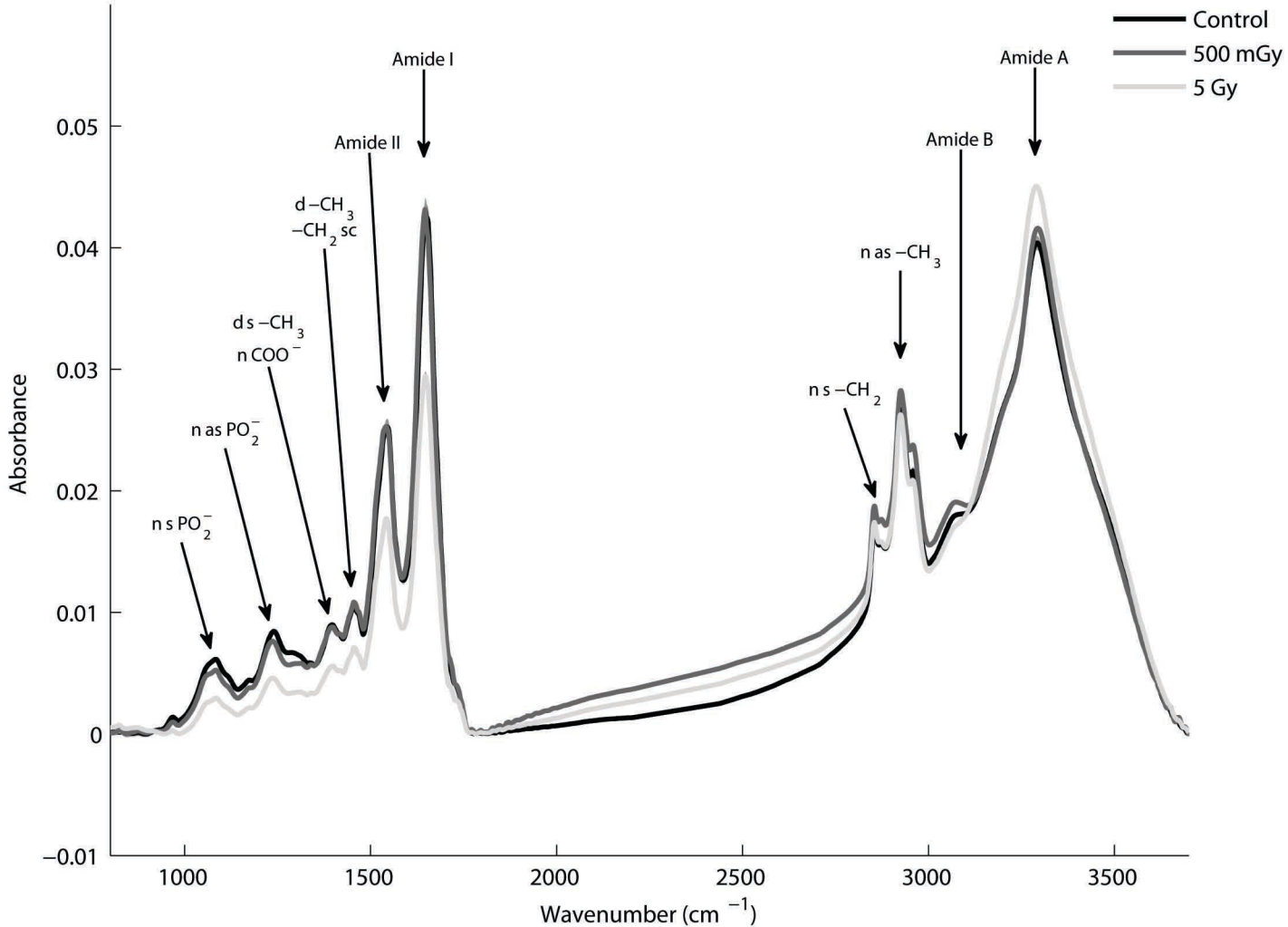
Table 3:

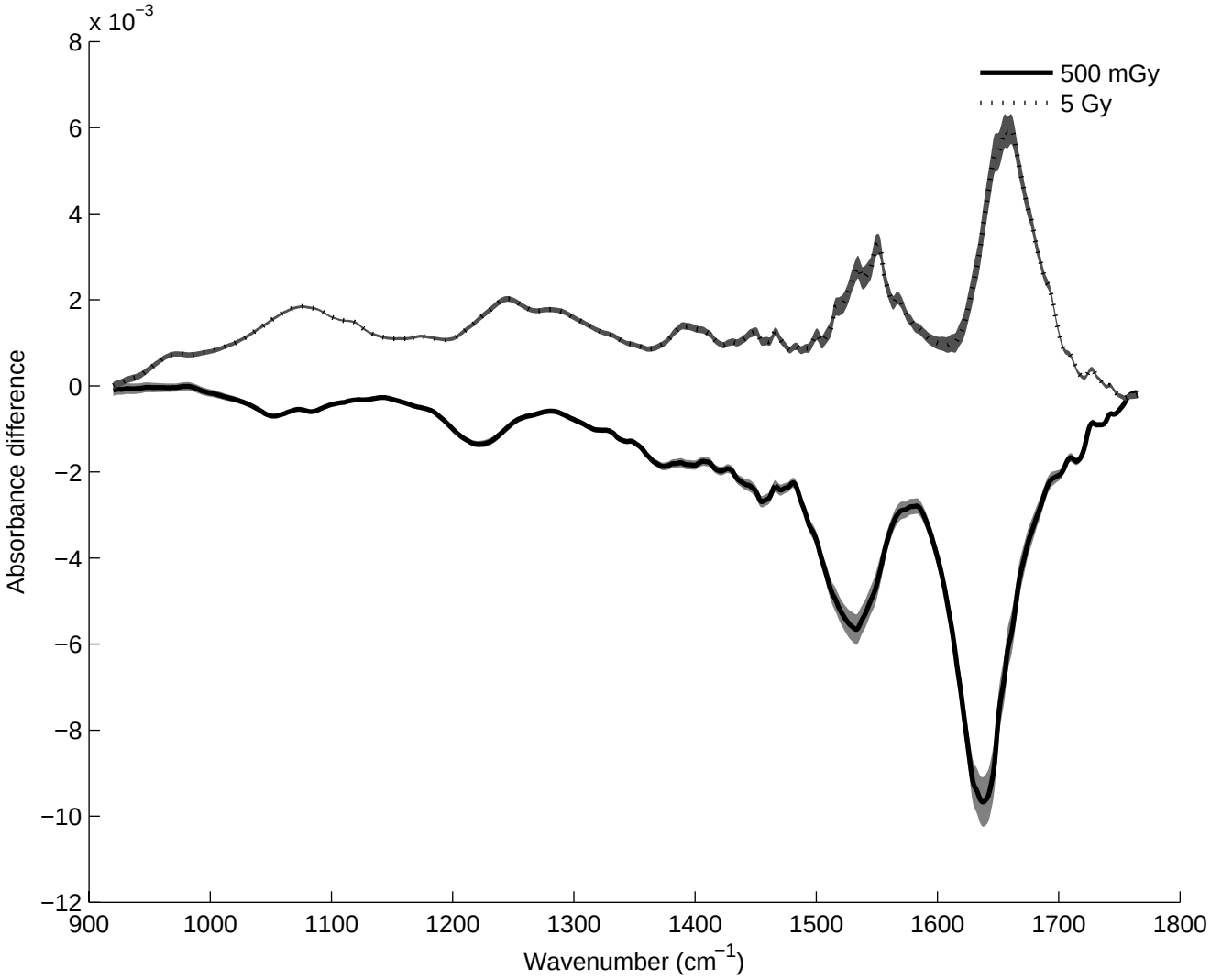
Time	nLv	RMSEC	RMSEP	R^2 Cal.	R^2 Test	Variance Described (LV 1)	Variance Described (LV 2)
Corrected Spectra							
6	22	0.35 (0.02)	0.48 (0.05)	0.97	0.94	96.6	2.4
12	21	0.56 (0.04)	0.76 (0.06)	0.93	0.86	97.2	1.9
24	21	0.27 (0.02)	0.40 (0.04)	0.98	0.96	97.4	2.2
48	24	0.32 (0.02)	0.46 (0.03)	0.98	0.95	95.4	3.7
96	20	0.39 (0.04)	0.52 (0.02)	0.96	0.93	94.6	1.7
1st Order Derivative							
6	12	0.29 (0.04)	0.39 (0.01)	0.98	0.96	94.3	2.6
12	15	0.48 (0.06)	0.87 (0.08)	0.95	0.82	96.1	1.9
24	18	0.17 (0.03)	0.50 (0.03)	0.99	0.94	96.9	0.8
48	24	0.16 (0.05)	0.58 (0.03)	0.99	0.91	95.2	2.7
96	16	0.26 (0.02)	0.56 (0.02)	0.98	0.91	95.7	1.7
2nd Order Derivative							
6	11	0.37 (0.03)	0.59 (0.05)	0.96	0.91	72.5	15.1
12	11	0.56 (0.07)	0.97 (0.05)	0.93	0.68	69.2	14.0
24	10	0.50 (0.05)	0.76 (0.05)	0.94	0.83	80.6	5.5
48	14	0.39 (0.14)	0.77 (0.07)	0.95	0.80	88.5	4.2
96	10	0.41 (0.13)	0.72 (0.06)	0.94	0.80	64.7	16.5

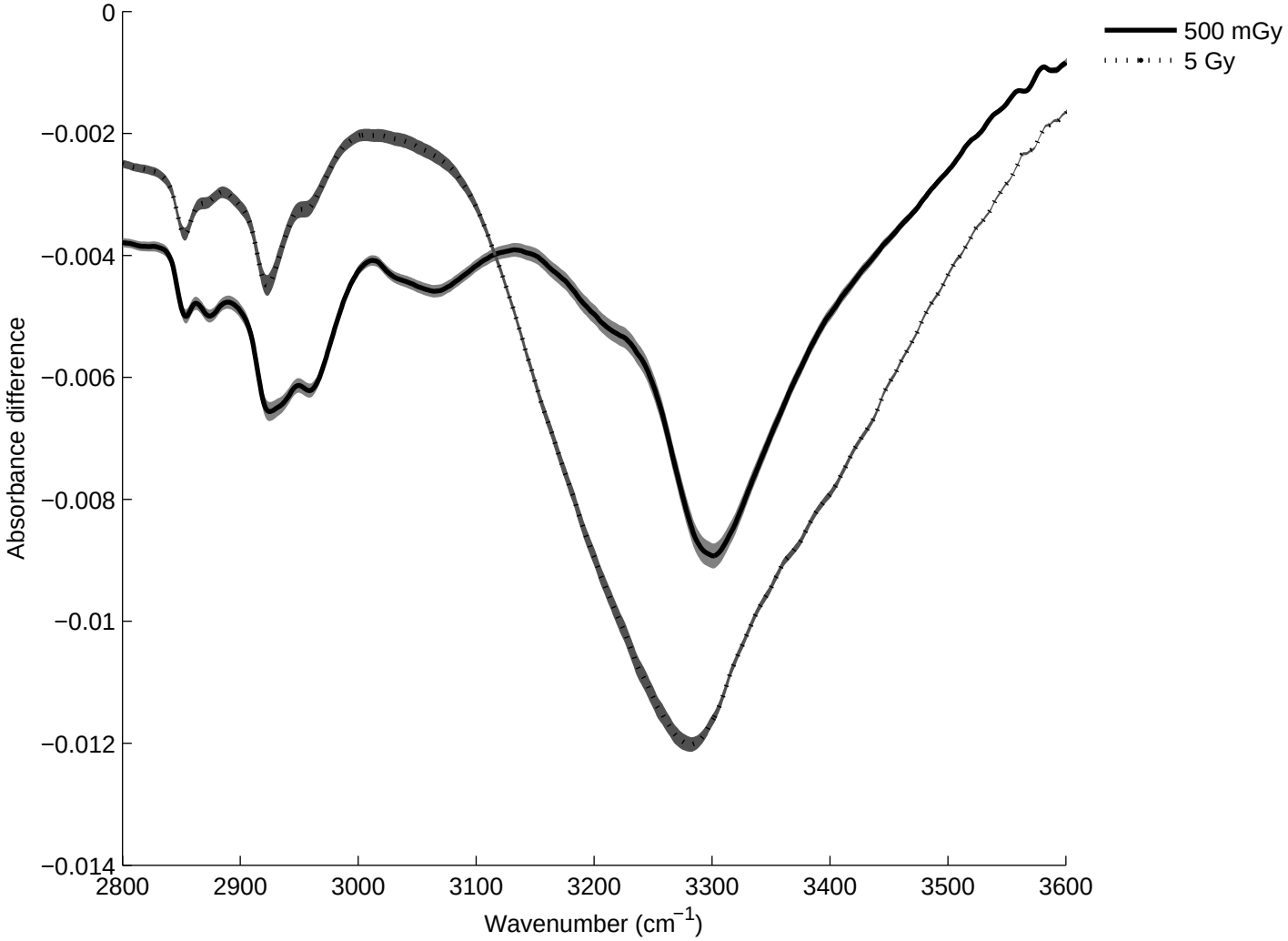
Table 4:

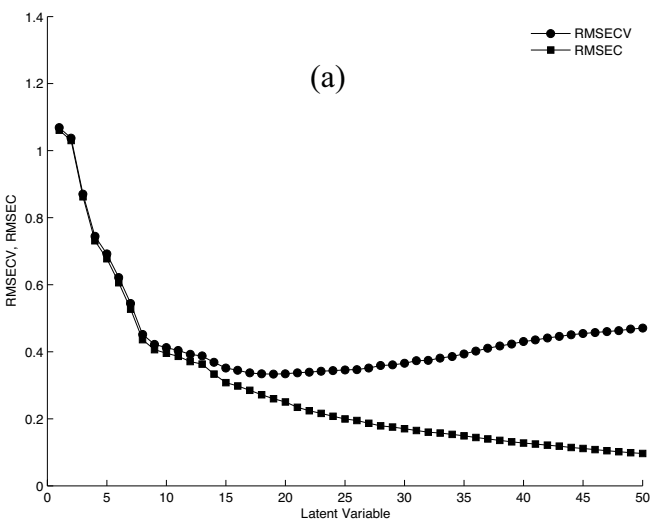
6 hr GRNN	**RMSEC**	**RMSEP**	**Smoothing**	**R^2 Cal**	**R^2 Test**
	0.028 (0.005)	0.042 (0.007)	0.018	0.999	0.999
12 hr GRNN	**RMSEC**	**RMSEP**	**Smoothing**	**R^2 Cal**	**R^2 Test**
	0.170 (0.007)	0.405 (0.072)	0.018	0.995	0.972
24 hr GRNN	**RMSEC**	**RMSEP**	**Smoothing**	**R^2 Cal**	**R^2 Test**
	0.094 (0.006)	0.171 (0.024)	0.019	0.998	0.994
48 hr GRNN	**RMSEC**	**RMSEP**	**Smoothing**	**R^2 Cal**	**R^2 Test**
	0.091 (0.062)	0.242 (0.082)	0.019	0.996	0.983
96 hr GRNN	**RMSEC**	**RMSEP**	**Smoothing**	**R^2 Cal**	**R^2 Test**
	0.003 (0.005)	0.005 (0.004)	0.014	0.999	0.999

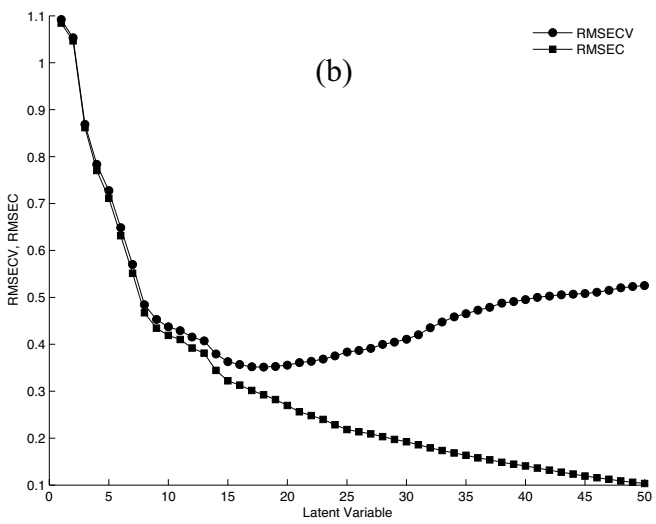

Input Layer
Radial Basis Function Layer (Hidden)
Summing Layer (Hidden)
Output Layer
X1
X2
X3
Xn
Y

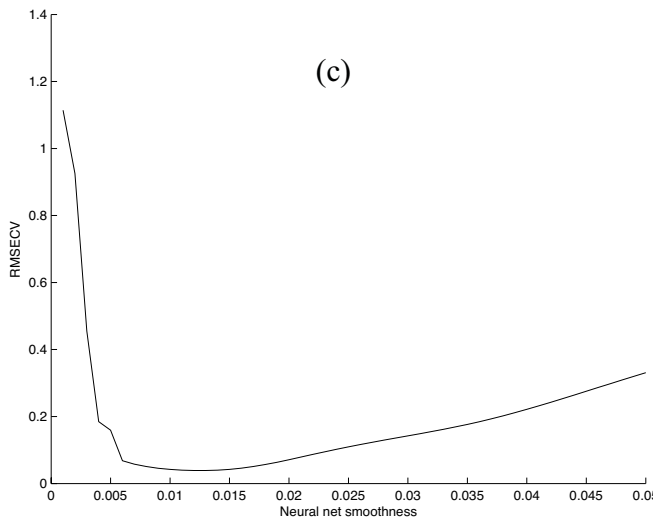

Control
500 mGy
5 Gy
Amide I
Amide II
Amide A
Amide B
d –CH3
–CH2 sc
d s –CH3
n COO−
n as PO2−
n s PO2−
n as –CH3
n s –CH2
Absorbance
0.05
0.04
0.03
0.02
0.01
0
−0.01
1000
1500
2000
2500
3000
3500
Wavenumber (cm −1)

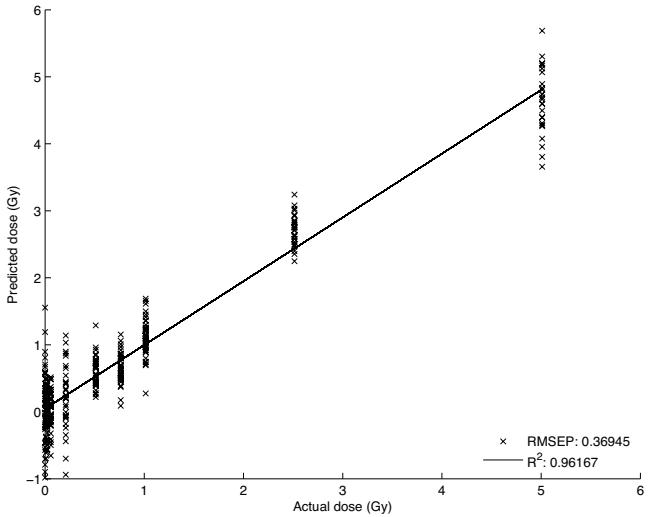

x 10^−3
8
6
4
2
0
−2
−4
−6
−8
−10
−12
Absorbance difference
500 mGy
5 Gy
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
Wavenumber (cm^−1)

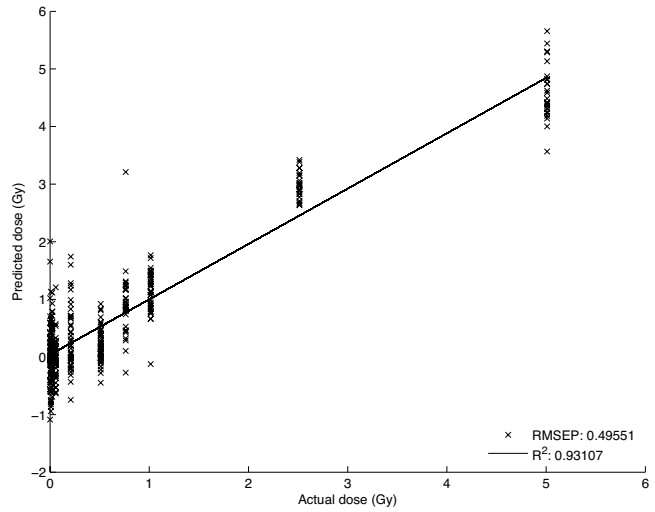

0
−0.002
−0.004
−0.006
−0.008
−0.01
−0.012
−0.014
Absorbance difference
2800
2900
3000
3100
3200
3300
3400
3500
3600
Wavenumber (cm^{-1})
500 mGy
5 Gy

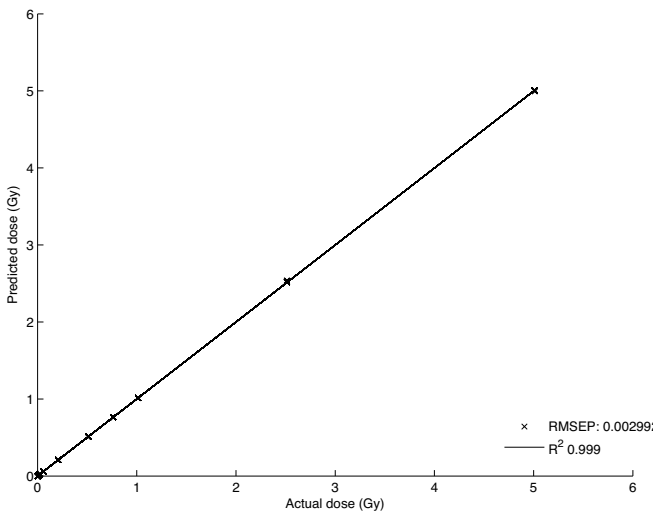

(a)
RMSECV
RMSEC
RMSECV, RMSEC
Latent Variable


(b)
RMSECV
RMSEC
RMSECV, RMSEC
Latent Variable


(c)
RMSECV
Neural net smoothness


Predicted dose (Gy)
RMSEP: 0.36945
R2: 0.96167
Actual dose (Gy)


Predicted dose (Gy)
RMSEP: 0.49551
R2: 0.93107
Actual dose (Gy)


Predicted dose (Gy)
R2 0.999
Actual dose (Gy)