From Positron to Pattern

Meles, Sanne K.; Kok, Jelmer G.; Renken, Remco J.; Leenders, Klaus L.

From Positron to Pattern: A Conceptual and Practical Overview of ¹⁸F-FDG PET Imaging and Spatial Covariance Analysis

4

Sanne K. Meles, Jelmer G. Kok, Remco J. Renken, and Klaus L. Leenders

Contents

S. K. Meles (✉)
Department of Neurology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
e-mail: s.k.meles@umcg.nl

J. G. Kok · R. J. Renken
Neuroimaging Center, Department of Neuroscience, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
e-mail: j.g.kok02@umcg.nl; r.j.renken@umcg.nl

K. L. Leenders
Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
e-mail: k.l.leenders@umcg.nl

Abstract

Imaging of brain glucose metabolism with ^{18}F-2-fluoro-2-deoxy-D-glucose posi-tron emission tomography (^{18}F-FDG PET) can give important information regarding disease-related changes in underlying neuronal systems, when com-bined with appropriate analytical methods. One such method is the scaled sub-profile model combined with principal component analysis (SSM PCA). This model takes into account the relationships (covariance) between voxels to iden-tify disease-related patterns. By quantifying disease-related pattern expression on a scan-by-scan basis, this technique allows objective assessment of disease activity in individual subjects. This chapter provides an overview of steps involved in pattern identification in ^{18}F-FDG PET data and is divided into three sections. Section 1 introduces basic concepts in nuclear imaging and explores the cellular underpinnings of signals measured with ^{18}F-FDG PET. Section 2 describes relevant basic concepts in ^{18}F-FDG PET image analysis including ana-tomical registration, normalization, and analysis of variance and covariance. Section 3 is dedicated to SSM PCA specifically. The goal of this chapter is to make the technique more accessible to readers without a mathematics or neuro-imaging background. Although many excellent texts on this topic exist, the current chapter aims to provide a more conceptual overview, including some discussion points that are not always formally described in literature.

4.1 ^{18}F-FDG PET Imaging

4.1.1   Basic Concepts in PET

Positron emission tomography (PET) allows measurement of the local tissue accu-mulations of injected radioactive tracers. The type of tracer that is used depends on the focus of the PET study. For instance, tracers can bind to specific receptor sites, allowing quantification of the distribution of a receptor in a tissue. Tracers can also be metabolically active compounds and allow measurement of the activity of a particular enzyme or biochemical pathway. Tracers are "tagged" with a radioactive atom. Radioactive decay of this atom is central to PET technology. A simplified explanation of radioactive decay necessary to understand PET technology is pro-vided in the next paragraphs.

Atoms consist of protons, neutrons, and electrons. Protons and neutrons can be found in the nucleus, whereas electrons orbit around the nucleus. Protons have a

positive charge, and electrons are negatively charged. The number of protons determines to what chemical element the atom belongs. Elements can have multiple isotopes. Isotopes of an element have the same number of protons in their nucleus, but a variable number of neutrons. In order for a nucleus to be stable, a certain balance is needed between protons and neutrons in the nucleus. Most naturally occurring isotopes have stable nuclei. Isotopes of an element with an unstable nucleus are referred to as radioisotopes. These isotopes will spontaneously emit particles or photons (or both) from its nucleus in order to regain stability. In this process, mass is converted into energy. This is called radioactive decay.

Different modes of radioactive decay exist. PET is designed to measure positron emission. In radioactive decay by positron emission, a proton in the nucleus is transformed into a neutron and a positron. A positron is the antiparticle of an electron (i.e., a positively charged electron, also called a β+ particle). When a positron is emitted, it travels a distance before it annihilates with an electron (a β− particle) from the surrounding matter. The annihilation of the masses of the two β particles results in the conversion and emission of two gamma (γ) rays. Gamma rays consist of high-energy photons. In the case of positron emission and annihilation, each γ-ray contains 511 keV in energy. These two γ-rays always originate simultaneously and are emitted in opposite directions. They form a so-called back-to-back pair (Fig. 4.1a).
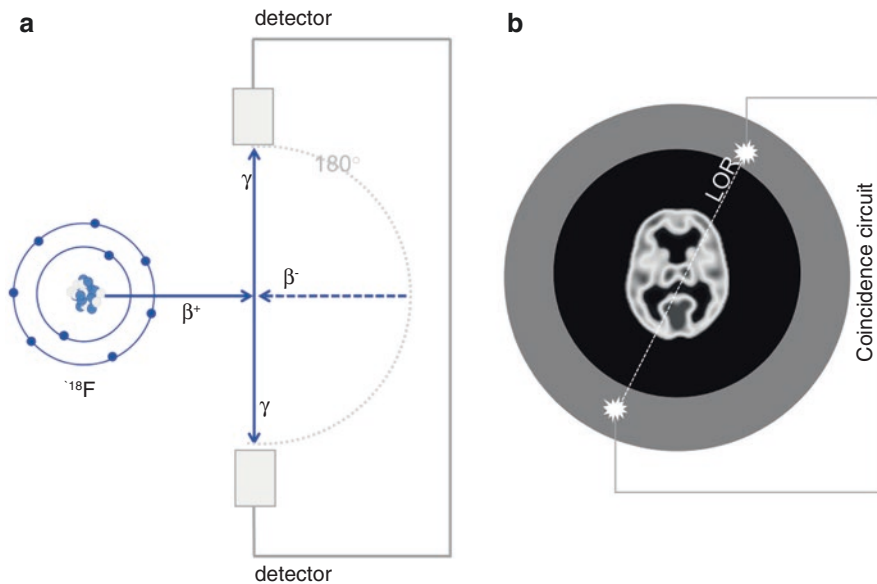
Fig. 4.1 Schematic of an annihilation event in positron emission (**a**) and its detection (**b**). (**a**)The positron travels a short distance before it loses its kinetic energy and then annihilates with an electron (β$^-$) from the surrounding matter. The mass of the two particles is converted into two opposing photon beams (γ-rays), traveling at approximately 180° from each other with an energy of 511 keV each. (**b**) In a PET camera, a ring of detectors is placed around the patient. Opposing detectors, connected via a coincidence circuit, record annihilation photons only when they arrive simultaneously. The origin of the annihilation event is inferred along the line of response (LOR)

A PET camera consists of a large number of small scintillation detectors positioned on a ring around the patient. Scintillation detectors use inorganic crystals that absorb γ-rays and then fluoresce. The γ-rays are converted into visible light by the crystal. The light signal is multiplied and transformed into an electric current. The strength of this current is proportional to the intensity of the light from the crystal and thus to the energy of the γ-ray that was detected.

A PET camera can determine the origin of the detected γ-ray in the tissue of the patient, by using the principle of γ-ray pairs. If one γ-ray is detected by a crystal, then its twin must be detected by the opposite crystal within a certain time window (a few nanoseconds). Such an event is called a coincidence event. Opposing crystals are linked via coincidence circuits. From a coincidence event, the origin of the annihilation can be inferred along a "line of response" (LOR) running between the two detectors (Fig. 4.1b). Modern electronics permit measurement of the time interval between detection of the first and the second photon of the same photon pair. This means that the origin of the photon pairs (the point of annihilation) can be pinpointed to a part of the LOR, close to the true event (Karp et al. 2008; Surti and Karp 2016). PET data thus consists of many back-to-back photon pairs, connected with LORs through the patient, from which the locations of the multiple annihilations are estimated. Because the detector system is a stationary ring that completely surrounds the patients, it is possible to acquire data from many different angular views (projections) simultaneously. From these multiple projections of the detected emissions, images can be reconstructed using mathematical algorithms (Cherry et al. 2012). In the final 3D PET image, each pixel (or voxel) has a value which reflects the number of coincidence events ("counts") that belong to that particular coordinate. Thus, the more radioactive decay in a certain part of the tissue, the higher the counts for the corresponding pixel in the image.

The spatial resolution of PET depends on its accuracy and precision in pinpointing the exact location of annihilation events. Even with sophisticated computerized techniques, the reconstructed location of the annihilation event is not exact. Current state-of-the-art PET systems have a maximum resolution of just under 3 mm. This limitation is inherent to physical properties of PET. First, the PET system assumes γ-rays pairs to be emitted at a 180° angle, but this is not always the case. Second, after emission, the positron travels a short distance (a few millimeters (Phelps et al. 1975)) before it annihilates with an electron.

Some commonly used PET tracers in neurology and their half-lives are listed in Table 4.1. Fluorine-18 (^{18}F) is the most commonly used radioisotope in clinical

Table 4.1 Commonly used isotopes in PET

| Isotope	Half-life (min)	Product	Examples of tracers
^{11}C	20.38	^{11}B	^{11}C-methionine: amino-acid transport ^{11}C-raclopride: dopamine receptors
^{15}O	2.03	^{15}N	^{15}O-water: perfusion ^{15}O-oxygen: oxygen utilization
^{18}F	109.8	^{18}O	^{18}F-FDG: glucose utilization ^{18}F-Fdopa: activity of aromatic amino-acid decarboxylase

practice. It is produced in a particle accelerator (cyclotron). An important advantage of ^{18}F is its relatively long half-life, which facilitates regional production of ^{18}F tracers and distribution to other hospitals. The main application of ^{18}F is labeling of fluorodeoxyglucose (FDG), which provides a measure of glucose utilization in the cells of the body. ^{18}F-FDG is the most widely used positron-emitting radiopharmaceutical with a wide range of clinical applications.

4.1.2 ^{18}F-FDG PET Imaging

The tracer ^{18}F-2-fluoro-2-deoxy-D-glucose (^{18}F-FDG) is an analogue of glucose. In normal conditions, glucose is the predominant metabolic substrate for brain tissue, and the brain's oxygen consumption is almost entirely for the oxidative metabolism of glucose (Fox et al. 1988). The brain holds minimal glycogen stores, and therefore, a permanent supply of glucose via the blood is necessary. Glucose is transported through the blood-brain barrier via GLUT transporters. Once in the cell, glucose undergoes numerous transformations to end up in three main metabolic pathways. The goal of each of these pathways is to create energy (in the form of ATP) for cells to function. The first step for any of these pathways is the phosphorylation of glucose into glucose-6-phosphate, a reaction catalyzed by the enzyme hexokinase. Hexokinase is the rate-controlling enzyme for all of the subsequent pathways. The enzymatic rate of this first step is equivalent to measuring the glucose utilization rate (Fig. 4.2a).

It has been attempted to quantify regional glucose metabolism with ^{14}C-glucose, but the many transformations and pathways for radioactively labeled glucose to enter are complex, leading to many different metabolites. Moreover, ^{14}C-glucose is very rapidly converted to CO_2 and H_2O, and CO_2 is too rapidly cleared from the cerebral tissue to allow measurement (Raichle et al. 1975; Sacks 1957). In the 1970s, this problem was solved by Sokoloff and colleagues, who applied a deoxyglucose analogue, 2-deoxy-D-glucose (2-DG), labeled with ^{14}C (Sokoloff et al. 1977). The deoxy variant of glucose is phosphorylated by hexokinase, at a definable rate relative to that of glucose. However, unlike glucose-6-phosphate, ^{14}C-2-DG-6-phosphate is not metabolized further and is essentially trapped in the tissue, allowing quantification of hexokinase. Therefore, regional deoxyglucose uptake measured with PET reflects the first step of the glucose metabolic pathway. Reivich et al. were the first to measure cerebral glucose metabolism with the deoxyglucose method in humans (Reivich et al. 1979). Instead of ^{14}C, ^{18}F was used as the radioisotope. ^{18}F-FDG behaves similarly as ^{14}C-DG and glucose (because fluorine behaves biochemically like hydrogen) and can measure glucose utilization accurately and reliably.

The kinetics of accumulation of ^{18}F-FDG-6-PO$_4$ can be described with a three-compartment model (Reivich et al. 1979). A description of tracer kinetic modeling is beyond the scope of this chapter and can be found elsewhere (Heiss 2014). In brief, after intravenous administration of ^{18}F-FDG, the regional cerebral metabolic rate of glucose (CMRglc) can be determined using the ^{18}F concentration in the tissue (measured with PET), the concentration of ^{18}F-FDG in the arterial plasma
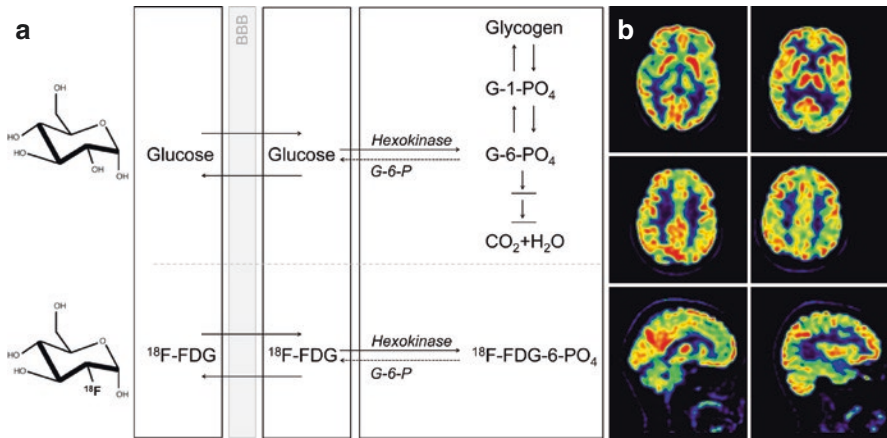
Fig. 4.2 (**a**) Schematic of the behavior of glucose and ^{18}F-FDG in brain tissue. Glucose and ^{18}F-FDG are similarly transported over the blood-brain barrier (BBB) and are both metabolites for the enzyme hexokinase. FDG-6-phosphate (^{18}F-FDG-6-PO$_4$) is trapped in the brain tissue, whereas glucose-6-phosphate (G-6-PO$_4$) can be metabolized further and has many metabolites. The dashed arrow represents the activity of glucose-6-phosphatase (G-6-P), which catalyzes the hydrolysis of G-6-PO$_4$ and ^{18}F-FDG-6-PO$_4$ back to glucose, and ^{18}F-FDG, respectively. This is a slow process (Sokoloff et al. 1977). (Adapted from (Heiss 2014) with permission from Springer Verlag Heidelberg). (**b**) Visual representation of an ^{18}F-FDG PET study of a 65-year-old healthy individual. The participant fasted for at least 6 h before the investigation. ^{18}F-FDG PET imaging was performed in a 3D mode using a Siemens Biograph mCT-64 PET/CT system. A 6 min static frame was acquired starting 30 min after the injection of 205 MBq ^{18}F-FDG in 4 mL saline. ^{18}F-FDG uptake and image acquisition were performed in the resting state with eyes closed in a dimly lit room with minimal auditory stimulation. The image was iteratively reconstructed with OSEM 3D, including point spread function and time-of-flight modeling (3 iterations/21 subsets, matrix 400), and smoothed with a Gaussian 2 mm full-width at half-maximum filter. Voxel size is 2 mm. Scatter and attenuation corrections were applied based on the acquired low-dose CT

(time-activity curve of blood tracer concentration), and the concentration of glucose in the plasma. In this situation, multiple sequential PET images are obtained (i.e., a dynamic protocol). When performed in this manner, PET provides absolute measures of regional CMRglc (i.e., in physiological units).

Thus, for a fully quantitative determination of *absolute* regional glucose utilization, arterial blood sampling is required, which is an invasive and time-consuming procedure. Already early on it was recognized that "raw counts" data could be analyzed instead of physiological units, obviating the need for arterial blood sampling (Fox et al. 1984). In current clinical practice and most experimental designs, arterial blood sampling is not strictly necessary, because the *relative* regional distribution of ^{18}F-FDG (raw counts) can be visually assessed and/or statistically analyzed.

In a clinical setting, ^{18}F-FDG is injected intravenously, and patients subsequently rest in a quiet, dimly lit room for 30–45 min, at which time metabolic equilibrium is reached. Next, a single static image with a frame duration of 5–15 min is acquired, and a low-dose CT scan is performed for attenuation correction. The corrected, reconstructed ^{18}F-FDG PET images are visually assessed by an expert reader in the context of the available clinical information. When performed according to the

guidelines, ^{18}F-FDG PET imaging in a clinical setting is easy, reliable, and accurate (Varrone et al. 2009). ^{18}F-FDG PET has an established role for a number of diagnostic indications in neurology, one of which is the differential diagnosis of neurodegenerative brain diseases.

An example of an ^{18}F-FDG PET image of a healthy control participant is given in Fig. 4.2b. Resting-state ^{18}F-FDG uptake is much higher in gray matter compared with white matter. In healthy controls, ^{18}F-FDG uptake is typically highest in the basal ganglia, primary visual cortex, cingulate cortex, and frontal cortex, with lower values in other cortical and subcortical areas, brain stem, and cerebellum (Heiss 2014).

4.1.3   Studying Brain Function with ^{18}F-FDG PET

The goal of neuroimaging is to understand how the brain functions under different circumstances and conditions, including disease. In the past decade, the focus of neuroimaging studies has converged on the study of brain networks (Friston 2011). Networks encompass connections between neurons and can be described in terms of structure and function. Structure dictates which neurons are connected. Function is dynamic, and this term is used to describe neuronal activity that assembles on the backbone of a relatively fixed anatomical structure (Buzsaki et al. 2013). A synapse may be present between two neurons, but the connection may be used to different degrees depending on the situation (Fornito et al. 2016). The main principle of functional neuroimaging techniques is that localized changes in neuronal activity can be mapped by measuring changes in energy metabolism or hemodynamics, which are thought to reflect the underlying cellular events.

Brain activity is determined by signaling between neurons. Neural signaling is achieved with the generation and propagation of action potentials across synapses. Energy metabolism (glucose metabolism) increases almost linearly with the frequency of action potentials (Kadekaro et al. 1985). Action potentials themselves do not require energy, as they are passive electrical consequences of K$^+$ and Na$^+$ fluxes across the cell membrane upon depolarization. Restoring ionic gradients and resting membrane potentials in the cell after an action potential does require energy. In primates, the major energetic burden is located at the nerve terminals from the postsynaptic neuron (the neuropil) (Sokoloff 1993). This is because there are a large number of synapses per neuron, and during signaling, postsynaptic passive Na$^+$ influx acts as an amplifier of the initial signal. Reversing ion fluxes after postsynaptic currents has been estimated to cost 74% of the energy used in signaling (Attwell and Laughlin 2001). Of note, energy metabolism in the postsynaptic neuron increases with both excitatory (glutamatergic) and inhibitory (GABA-ergic) signaling (Buzsaki et al. 2007; Jueptner and Weiller 1995). The only way to determine which has occurred is to look downstream at the next synapses in the projection zones of those neurons (Sokoloff 1993).

To sustain brain activity, neurons continuously require energy in the form of adenosine triphosphate (ATP). Under normal physiological conditions, generation of ATP is supported almost exclusively by the oxidative (i.e., aerobic) metabolism

of glucose. Only around 10% of ATP is generated by anaerobic metabolism through glycolysis (Vaishnavi et al. 2010). The products of glycolysis, such as lactate, can subsequently be metabolized further using oxygen. Astroglia are thought to play an important role in the latter process, in which lactate is also exchanged between neurons and glia cells (Magistretti and Allaman 2015).

Glucose and oxygen supplies are maintained by an adequate regulation of cerebral blood flow (CBF). The relationship of the cerebral metabolic rate of glucose (CMRglc), the cerebral metabolic rate of oxygen (CMRO$_2$), and cerebral blood flow (CBF) to the underlying cellular events is complex and still not completely understood. Counter-intuitively, CBF increases do not simply serve to adjust glucose and oxygen delivery to the variable energy demands of neuronal tissue. Activation studies have shown that in response to a task (i.e., an increase in neuronal activity), CBF and CMRglc increase together, but far exceed CMRO$_2$ (Fox and Raichle 1986; Fox et al. 1988). Several models have been used to explain the cellular underpinnings of neurovascular coupling and uncoupling (Lin et al. 2008; Lin et al. 2009; Lin et al. 2010).

An important advantage of measuring CMRglc is that it provides a direct, physiologically specific signal that can be quantified. This was elegantly demonstrated by several early autoradiography studies with ^{14}C-deoxyglucose (Sokoloff 1993). For example, in rats, retinal stimulation with flashes of light of a known, calibrated intensity resulted in proportional increases in local CMRglc in the primary projection areas from the retina, whereas local CMRglc remained unchanged in structures that did not receive direct projections from the retina (Batipps et al. 1981).[18]F-FDG PET measurements are indicative of a steady state of neuronal activity during the uptake and scanning interval.

Relative to other tissues, the brain's energy demand is high in the resting state and during sleep and increases with only a fraction of its baseline metabolism with activity. In the resting state, most of this energy is also devoted to neuronal signaling (i.e., synaptic function) (Sibson et al. 1998). In activation (or task-based) studies, local changes in ^{18}F-FDG uptake in response to a task are studied to localize brain functions. Studies in which subjects are in a resting state, which means there is no specific sensory stimulation and patients are not engaged in any behavioral or physical task, can give information about time-invariant aspects of brain function. Several neuroimaging and electroencephalography (EEG) studies have shown that spontaneous neuronal activity is highly organized at rest and that several conditions (including disease) can alter resting-state neuronal activity.

4.2   Analysis of Resting-State ^{18}F-FDG PET Images

4.2.1   Image Registration

Voxel-wise image analyses usually start with the registration of each image to standard space. This is because this type of analysis of brain images is hampered by the differences in brain morphology between subjects. In image registration, the brain
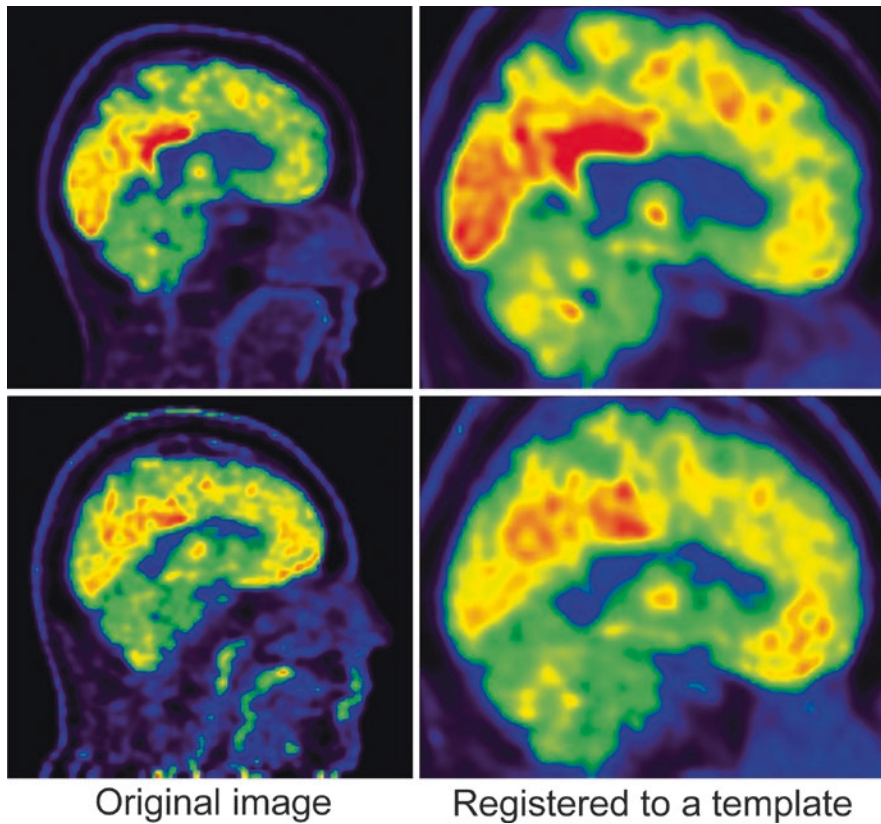
Fig. 4.3 Example of two healthy controls showing the reconstructed ¹⁸F-FDG PET image, and the images after registration to an ¹⁸F-FDG PET specific template in standard space (Della Rosa et al. 2014)

images of subjects are translated to another image (usually a template), such that voxels/regions can be compared between subjects. Image registration entails the estimation of the optimal spatial transformation between two images. ¹⁸F-FDG PET images are often directly registered to a template such that all the images are in the same space. An example of ¹⁸F-FDG PET image registration is given in Fig. 4.3. A more detailed explanation of image registration can be found elsewhere (Herholz et al. 2004).

4.2.2 Normalization

By adhering to strict scanning protocols, it is attempted to minimize the differences between each scanning session (Varrone et al. 2009). By image registration, morphological differences are accounted for. However, considerable inter-individual
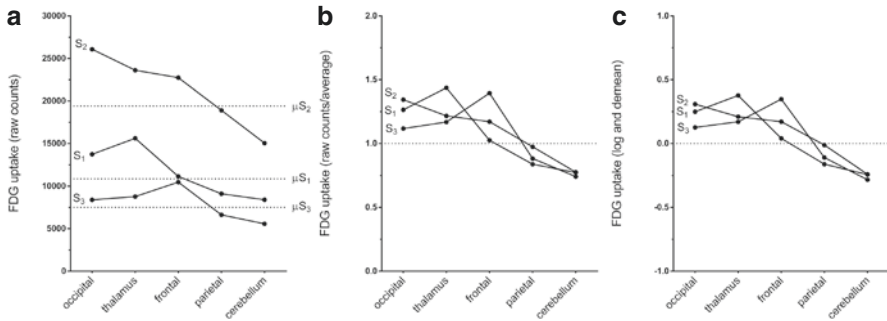
Fig. 4.4 Two types of normalization. Raw voxel values (i.e., raw count data) (**a**); ratio-normalized voxel values (**b**); and voxel values normalized according to the SSM (**c**) are depicted for five different coordinates in three healthy controls (S_1, S_2, and S_3). The average voxel value was calculated for each subject, by taking all voxels in the image within a gray matter mask. This average voxel value is indicated for each subject in (**a**) with the dashed line (μS_1, μS_2, and μS_3). In (**b**) each voxel value was divided by the corresponding subject average. This results in voxel values centered around 1 for each subject. In (**c**) voxel values within the same mask were log-transformed, and subsequently the log mean was subtracted from the data. This results in voxel values centered around zero for each coordinate and each subject. In (**a** and **b**), data is non-negative. In (**c**), negative values are present

differences in ^{18}F-FDG uptake remain present. This is because there may be slight differences in the dose of ^{18}F-FDG injected, or due to baseline metabolic differences between people (even after fasting). This is apparent in Fig. 4.4a, which shows the average brain FDG uptake for three healthy controls, scanned with identical protocols. These large inter-individual differences will obscure underlying task-evoked or disease-dependent patterns of altered metabolism.

This can be solved with tracer kinetic modeling, which transforms count data into physiological units using information from arterial blood sampling. However, as stated previously, this entails an invasive, time-consuming procedure. A solution is normalizing the raw count data to a reference value. It should be noted that any ^{18}F-FDG PET study that does not apply arterial blood sampling cannot study absolute differences of ^{18}F-FDG uptake, but can only make inferences on relative differences.

Several approaches are used to normalize raw count data. Some researchers choose a reference region which is thought to be unaffected by the disease process. The average ^{18}F-FDG uptake in that region is measured, and all voxel values in each image are subsequently divided by this value (Borghammer et al. 2008). An important limitation of this approach is that it requires a priori assumptions. For example, in the study of Parkinson's disease, some authors have chosen the cerebellum as a reference region, whereas it is now known that the cerebellum plays an important part in parkinsonism (Bostan et al. 2013; Rodriguez-Oroz et al. 2009). For ^{18}F-FDG studies which study brain-wide metabolism, choosing a reference region may therefore be problematic. Still, in other radiotracer studies, a reference region can be very useful.

A frequently used alternative is to ratio-normalize each voxel value to the subject's average whole-brain uptake (usually within a gray matter mask), which is

referred to as global mean normalization (Fig. 4.4b). An equivalent approach is the scaled subprofile model (SSM), in which the data is first log-transformed, and subsequently the log mean is subtracted (Fig. 4.4c). The appendix provides additional details on these two methods and the associated issues.

4.2.3   Analysis of Variance and Covariance

After applying some type of normalization, a straightforward approach to investigating differences in cerebral glucose metabolism between patients and controls is by comparing the mean ¹⁸F-FDG uptake in each region (or each voxel) between the groups with multiple *t* tests. This is an example of analysis of *variance*. Variance is a measure of spread of the data. It is equal to the square of the standard deviation of the data:

$$\text{var}(X) = s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}$$

where $\bar{X}$ indicates the average of variable *X*. This formula can be rewritten as:

$$\text{var}(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{(n-1)}$$

Multiple *t* tests between voxel values in two or more groups can be easily performed. If a cluster of voxels holds significantly lower values in patients compared to controls, then this brain region shows decreased ¹⁸F-FDG uptake in the disease state, which reflects a loss of synaptic integrity. The interpretation of such statistical parametric mapping (SPM)-based group contrasts is thus very straightforward. Several studies have used a voxel-based SPM analysis to identify group differences between patients with a neurodegenerative disease and healthy age-matched controls (Eckert et al. 2005; Juh et al. 2004; Teune et al. 2010; Yong et al. 2007). These univariate patterns give a good impression of the brain regions involved in disease.

Covariance is a measure of how much two variables change together. The formula for covariance is very similar to the formula for variance, but includes an additional variable (variable *Y*):

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

A positive covariance indicates that as *X* increases, so does *Y*; and a negative covariance indicates that as *X* increases, *Y* decreases (or vice versa). If the covariance is zero, *X* and *Y* are unrelated. The concepts variance and covariance are explained in a schematic in Fig. 4.5.

If there are more than two variables (i.e., more than two dimensions), covariance (C) can be stored in a matrix. If we have three variables (*x*, *y*, and *z*), the diagonal
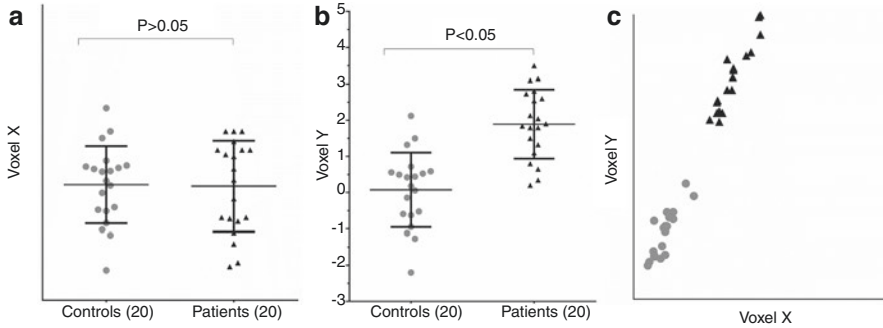
Fig. 4.5 A schematic, easy representation of variance and covariance (mock data). Two voxels are considered in a patient group and a control group. The mean in ^{18}F-FDG uptake (normalized counts) in voxel X is compared between patients and controls with a t test (**a**), which does not show a significant difference. The mean value for voxel Y does show a significant difference (patients higher than controls) (**b**). Voxel X and voxel Y are correlated; a higher value in voxel X predicts a higher value in voxel Y. These two variables thus show positive covariance (C)

entries in this matrix reflect the variance. The off-diagonal entries reflect the covariance between x and y, x and z, and y and z. The matrix is symmetric around the diagonal.

$$C = \begin{pmatrix} \mathbf{cov}(\mathbf{x,x}) & \mathrm{cov}(x,y) & \mathrm{cov}(x,z) \\ \mathrm{cov}(y,x) & \mathbf{cov}(\mathbf{y,y}) & \mathrm{cov}(y,z) \\ \mathrm{cov}(z,x) & \mathrm{cov}(z,y) & \mathbf{cov}(\mathbf{z,z}) \end{pmatrix}$$

4.2.4 Principal Component Analysis

In data with only a few dimensions, it is easy to appreciate the "patterns" in the data (i.e., the relationship between variables). However, when the dimensionality of the data increases, as is the case in typical neuroimaging data (>100,000 voxels and dozens of subjects), relationships between variables can no longer be presented graphically. Principal component analysis (PCA) reduces the number of dimensions and can hereby aid in identifying patterns in complex datasets. In this section, we will explore what PCA does in a simple, two-dimensional example.

Imagine we have studied the values of two voxels (X and Y) in ten controls and ten patients. PCA requires a dataset with a mean of zero. Therefore, we first subtract $\bar{X}$ from each observation of X and $\bar{Y}$ from each observation of Y. A plot of the data is shown in Fig. 4.6 (this is mock data, so the values do not represent true voxel values).

From Fig. 4.6b, it is clear that the two voxels are related. We can infer that these two variables have a positive covariance. The variance-covariance matrix for these variables is:

a

	Voxel X	**Voxel Y**
Subject 1	-3.58	-1.74
Subject 2	-2.68	-0.44
…	…	…
Subject 20	5.02	1.66
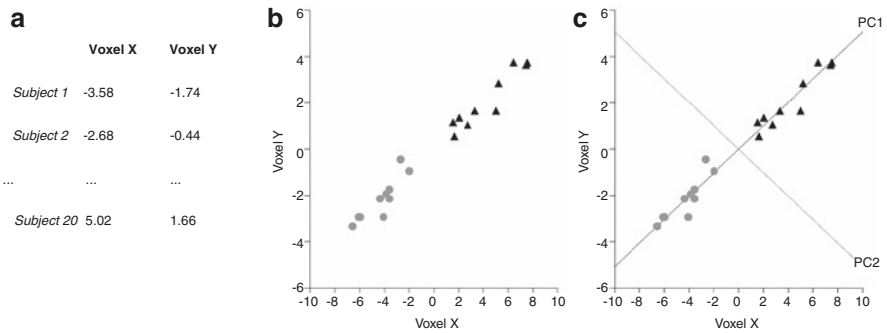
Fig. 4.6 Two voxels were studied in 20 subjects (mock data). The data points were demeaned (**a**). (**b**) shows these values for voxel *X* and voxel *Y* in ten controls (grey circles) and ten patients (black triangles). In (**c**), the first eigenvector (PC1) and the second eigenvector (PC2) are drawn. PC2 is perpendicular (orthogonal) to PC1. PC = principal component

$$\text{cov}(X,Y) = \begin{matrix} 22.97 & 11.57 \\ 11.57 & 6.01 \end{matrix}$$

The variance of voxel *X* is 22.97 and the variance of voxel *Y* is 6.01. The covariance of the two voxels is 11.57. From the covariance matrix, we can calculate the **eigenvectors** and the **eigenvalues** of this dataset (how this is done is beyond the scope of this chapter):

$$eigenvectors = \begin{matrix} 0.45 & -0.89 \\ -0.89 & -0.45 \end{matrix} \qquad eigenvalues = \begin{matrix} 0.14 \\ 28.84 \end{matrix}$$

The eigenvectors describe the lines that are plotted in Fig. 4.6c. The first eigenvector, called principal component (PC) 1, almost perfectly fits the data points. The second eigenvector, PC2, describes how much the data points deviate from PC1. Thus, this process of taking the eigenvectors from the covariance matrix has enabled the extraction of lines that characterize the data.

The first eigenvector explains most of the variability in the data. This eigenvector therefore has the highest eigenvalue. In PCA, eigenvectors (or components) are always ordered in terms of how much of the variability in the data they describe. The component with the highest eigenvalue is principal component 1, the one with the second highest eigenvalue is principal component 2, and so on. The number of principal components depends on the dimensionality of the data. A PC is always perpendicular (also called *orthogonal*) to all other PCs.

The rest of the steps involve transforming the data such that they are expressed in terms of PC1 and PC2. This means that each data point will obtain a new value in terms of PC1 and PC2. This is essentially a rotation by which PC1 and PC2 describe the new axes. Note that in Fig. 4.7a, this rotation effectively removes the
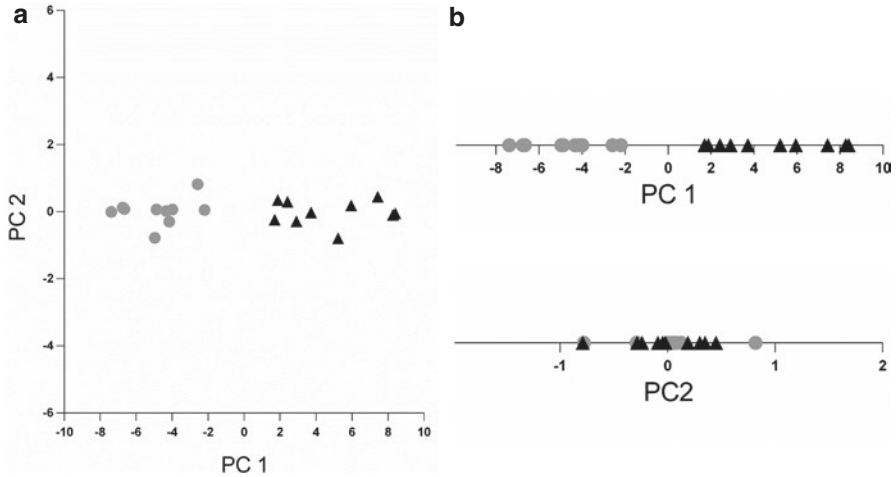
Fig. 4.7 Rotation of the data such that PC1 values are plotted along the *x*-axis, and PC2 values are plotted along the *y*-axis (**a**). The data can be reduced in dimensionality (from two-dimensional to one-dimensional) when only one PC is considered (**b**). PC1 contains the information of interest (it separates the two groups), and PC2 can be discarded

covariance from the data. The new values can be calculated by multiplying the original, demeaned voxel values with each eigenvector. We could also decide to only keep PC1, since this eigenvector describes the most important effects in the data, and we can discard PC2. This is a way of reducing the dimensionality of the data (Fig. 4.7b). We have discarded some of the information (the PC with the lowest eigenvalue), but kept the most important effects in one dimension (the PC with the largest eigenvalue).

In the example above, we have explained PCA with 20 subjects but just 2 voxels. In that mock data, there were only two eigenvectors. In neuroimaging studies, there are many more voxels than there are subjects. In such studies, the number of possible eigenvectors is limited to the number of subjects minus 1. This can be intuitively understood when we imagine a situation where we have just two subjects with three voxels each (voxels *x*, *y*, and *z*). The voxels determine the axes: a three-dimensional grid with axes *x*, *y*, and *z*. The subjects are plotted in this three-dimensional grid: each subject has a value for each voxel. With only two subjects, just one eigenvector (running exactly between the two points) can be calculated. Now imagine we have three subjects in the same space. In this situation, two eigenvectors can be calculated. Since these three subjects are in one plane, a third eigenvector cannot be determined. This same principle works for multi-dimensional space. In other words, the possible number of eigenvectors depends on the length of the shortest dimension. As noted above, in neuroimaging, the shortest dimension is usually the number of subjects. For example, in the typical situation where we have 40 subjects (20 controls and 20 patients) and > 100,000 voxels, 39 eigenvectors can be determined.

4.3   SSM PCA

Now that the concepts variance, covariance, and PCA have been explained, we will give an overview of the steps involved in scaled subprofile model (SSM) and principal component analysis (PCA). SSM PCA was first introduced by Moeller and colleagues for region-of-interest (ROI) data (Eidelberg et al. 1994; Moeller et al. 1987; Moeller and Strother 1991) and was later extended to whole-brain voxel-wise analyses (Eidelberg 2009; Habeck et al. 2008; Ma et al. 2007; Spetsieris and Eidelberg 2011). This approach combines a type of normalization (SSM) with principal component analysis (PCA) to find patterns (components) that can potentially discriminate between neuroimaging data of two groups. Here, we explain the method in a conceptual manner, such that an audience with a limited mathematics background can follow the steps involved.

4.3.1   Defining the Data

Usually, SSM PCA is applied to ^{18}F-FDG PET data that are acquired in a static imaging protocol and registered to a template in Montreal Neurological Institute (MNI) brain space. The data are typically smoothed (8–10 mm full width at half maximum) to improve the signal to noise ratio. A threshold of the whole-brain maximum is applied to remove out-of-brain voxels. The threshold value is usually chosen as the value that is 35% of the whole-brain maximum. This results in a mask of mainly gray matter (see Spetsieris and Eidelberg (2011) for details and alternatives). This mask is created for each subject, and the masks for all subjects are combined to include only those voxels that are shared between all participants. The remaining data can be stored in a matrix where subjects are in rows and voxels are in columns (Fig. 4.8).

4.3.2   Normalization with the Scaled Subprofile Model (SSM)

First, the SSM is applied, which refers to the normalization of the raw count data. This preprocessing starts with the log transformation of each voxel value for each

Fig. 4.8 Data matrix (D) with m subjects (S_m) and n voxels (v_n). Removal of the subject mean (row-centering) and group mean (column-centering) is indicated
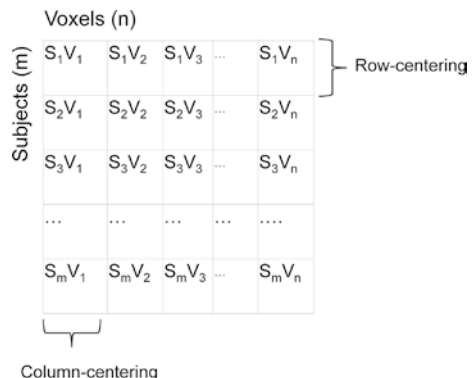
subject. Subsequently, the mean for each subject is removed (row-centering). These two steps combined serve to remove subject-specific scaling effects (also see the appendix and Fig. 4.4c). Next, the mean per voxel is removed (column-centering). This mean per voxel is referred to as the group mean profile (GMP). The remaining data matrix consists of residual voxel values for each subject. For each subject, this is termed the "subject residual profile" (SRP). Because all SRPs are in the same scale, the SRPs can now be compared between subjects (and thus between groups).

4.3.3 Calculating Eigenvectors from a Covariance Matrix

After normalization, a PCA is performed on the SRP data matrix. First, a covariance matrix is determined. The formula for covariance was introduced previously:

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)}$$

Remember that the SRP matrix has a mean of zero, because the data was log-transformed and the mean was subtracted. Therefore, we can ignore the terms $\overline{X}$ and $\overline{Y}$. In other words:

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i)(Y_i)}{(n-1)}$$

The covariance between all voxels across all subjects can be determined from the voxel * voxel covariance matrix. We arrive at the voxel * voxel covariance matrix by multiplying the transpose of D with D itself (D′ * D). This can be visualized as follows (Fig. 4.9):

Since we usually have thousands of voxels, the voxel * voxel data matrix (S_{vox}) will be quite large, and calculating the eigenvectors from S_{vox} will require a lot of computational power. This can be solved by determining the eigenvectors from the subject * subject covariance matrix (S_{sub}) instead and left-multiplying these with the transposed SRP (Spetsieris and Eidelberg 2011).

The eigenvectors from S_{vox} are referred to as group invariant subprofiles (GIS) in literature (Eidelberg 2009; Spetsieris and Eidelberg 2011). The terms eigenvector, GIS, and PC are often used interchangeably. As described in Sect. 2, the components in PCA are always ordered in terms of variance accounted for. Thus, PC1 explains most of the variance in the data, PC2 less, and so on. The last component accounts for only a small percentage of the total variance.

D'   **D**

Subjects (m)   Voxels (n)

S_1V_1	S_2V_1	S_3V_1	...	S_mV_1
S_1V_2	S_2V_2	S_3V_2	...	S_mV_2
S_1V_3	S_2V_3	S_3V_3	...	S_mV_3
...	...	...	...	....
S_1V_n	S_2V_n	S_3V_n	...	S_mV_n

Voxels (n)

*

Subjects (m)

S_1V_1	S_1V_2	S_1V_3	...	S_1V_n
S_2V_1	S_2V_2	S_2V_3	...	S_2V_n
S_3V_1	S_3V_2	S_3V_3	...	S_3V_n
...	...	...	...	....
S_mV_1	S_mV_2	S_mV_3	...	S_mV_n
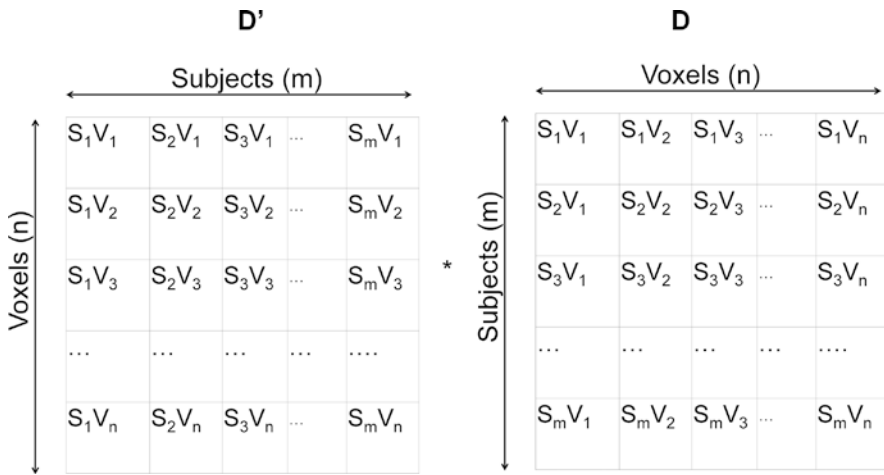
Fig. 4.9 The product (not shown here) will show the variance across the diagonal of the matrix, and the off-diagonal entries will show the covariance between the voxels. Thus, the entire matrix describes the spatial variance and covariance relationships between voxels across subjects

4.3.4   Calculating Subject Scores and Selecting Disease-Related Components

In order to reduce the dimensionality of the data, we need to project each data point (each SRP) onto each PC. Each subject will receive a score on each PC. These so-called subject scores (SS) are easily computed by the inner product of the two vectors:

$$SS = SRP * PC$$

In the case of n subjects, this will result in a matrix of subject scores on each PC, with a maximum of $n-1$ PCs (Fig. 4.10):

The subject scores will determine which PC is of interest to the study. On some PCs, the subject scores will be significantly different between patients and controls. These components may thus contain the disease-related changes that are of interest and may be selected for further analysis. Of note, the subject scores on each PC are inspected, to ensure that the mean subject score in patients is higher than the mean subject score in controls. This convention aids in the interpretation of the PC maps, as will be discussed later. In case the calculated mean scores are higher in controls than in patients, both the subject scores and the associated PC are multiplied by −1.

There are several ways to decide which PC (or combination of PCs) constitute(s) the final disease-related pattern (Spetsieris and Eidelberg 2011). In some studies, a

Fig. 4.10 Matrix of subject scores on each PC

single PC is chosen (usually PC1) if it discriminates significantly between controls and patients. Consecutive, smaller PCs are not included (even if these also discriminate between patients and controls) (Niethammer and Eidelberg 2012; Wu et al. 2013; Wu et al. 2014). A disadvantage of that approach is that it assumes that the relevant disease-related information is captured in a single component, which may not be the case. On the other hand, an important consideration is the risk of overfitting. Including more components may yield a pattern that gives a better fit of the initial sample, but may be limited in its relevance or generality across new datasets from the same population.

In previous studies by our group (Meles et al. 2018a; Meles et al. 2018b; Teune et al. 2013; Teune et al. 2014a; Teune et al. 2014b), combinations of principal components were selected using a forward stepwise logistic regression model. First, the components that explain the top 50% of the total variance in the data are selected. This is an arbitrary threshold that assumes that the lower 50% includes only noisy components that explain very small sources of variance in the data (a few percent) and are probably not disease-related. The combination of components that together give the lowest Akaike information criterion (AIC) of the model is selected (Akaike 1974). In other words, we combine the least possible number of components that together give the optimum discrimination between groups (trade-off between discriminative power and parsimony of the model). The selected components are subsequently combined linearly into a single PC vector using the coefficients determined by the logistic regression model.

We have re-evaluated the data in previously published disease-related patterns (Teune et al. 2013; Teune et al. 2014a; Teune et al. 2014b) and found that a combination of components as selected by the logistic model gave better discrimination of groups compared to the selection of PC1 alone. This does not imply that this model can be applied blindly to any dataset. In every analysis, each separate component should be inspected carefully, for instance, by visually checking if the component could potentially reflect disease activity or noise.

An important consideration is that PCA is susceptible to outliers. One outlier may contribute overwhelmingly to the variance, resulting in a first PC that accounts for most of the variance (i.e., >90%). The rest of the components are always orthogonal to this first, "faulty" PC, and thus even the remaining PCs are influenced by this issue, even though they reflect the effects of interest in the data (Habeck et al.

2010). It is thus advisable to always check each PC visually and in terms of variance accounted for. It may be necessary to re-run the analysis, excluding a problematic case.

Of note, when PCA is performed without this double-centering procedure, the chance that the first few components will reflect group-dependent differences in brain function is reduced (Moeller and Strother 1991; Spetsieris and Eidelberg 2011). In such a scenario, the first PC will reflect major sources of variance stemming from global mean values. This PC does not discriminate patients from controls. The disease-related pattern (i.e., the component that can discriminate between groups) shifts in order of importance, to a lower eigenvalue.

4.3.5   Prospective Application of the Pattern

Once a pattern has been identified, it can be applied to new scans. Scans of new subjects are registered to the same template and masked using the same parameters as in the original dataset. Next, the data are log-transformed, the mean per subject is subtracted, and the SRP for each subject is calculated by subtracting the group mean profile (GMP) that was determined from the original pattern identification dataset. Finally, the pattern is projected onto the new data to calculate the subject score:

$$SS = SRP * PC$$

Subject scores are usually z-transformed with reference to the control group:

$$Z_{SS} = \frac{SS - \mu_{SSHC}}{\sigma_{SSHC}}$$

SS refers to the ("raw") subject score, $\mu_{SS\,HC}$ refers to the mean raw subject score of the control group, $\sigma_{SS\,HC}$ refers to the standard deviation of the raw subject scores in the control group, and Z_{SS} refers to the z-transformed subject score of the new subject. This implies that the mean z-score of controls will be set to zero, with a standard deviation of 1. If the new dataset was acquired in a different manner than the original pattern identification dataset (for instance, the subjects were scanned on a different PET system), it may be necessary to z-transform the new data to a control cohort that was acquired under the same circumstances (Kogan et al. 2019).

4.3.6   Validation

One cannot assume that components that result from PCA are intrinsically meaningful. PCA is a mathematical operation, and aspects such as variance ordering and orthogonality are true by design (Habeck and Moeller 2011). Even if the signals only consist of noise, PCA will achieve a data reduction. Disease-related patterns

may be found by coincidence, and therefore, it is important to check empirically if the findings that were determined on the derivation set hold in a completely new dataset. If the pattern successfully discriminates between patients and controls in a new dataset, it is likely that this pattern can be interpreted as disease-related. This can be substantiated if subject scores correlate to other markers of disease such as disease duration or severity (measured with a validated scale) in new subjects. If significant correlations are found, this would mean that the identified component has some meaning other than a purely mathematical dimension.

In some cases, a testing set may not be available. This can be (partly) solved using a leave-one-out cross validation (LOOCV). In LOOCV, the analysis described above is repeated several times, each time leaving out one subject. Imagine patient *x* is part of our identification sample. We have included 20 controls and 20 patients. We leave out patient *x* and re-determine the pattern on the remaining 20 controls and 19 patients. We subsequently calculate the subject score of this pattern in patient *x*. We now have the LOOCV subject score for patient *x*. This procedure can be repeated for each subject (patients and controls). It results in subject scores for each subject, which are independent from the pattern identification step. The difference between LOOCV subject scores in controls and patients can be determined with a *t* test. If significant, the original pattern is considered a predictor for the disease in new cases.

4.3.7   Visualization and Interpretation of PC Maps

PCs can be visualized as 3D brain images, in which each voxel has a weight (the voxel value). Positive and negative weights indicate the direction of the principal component vector with respect to the mean. Voxels with a greater absolute weight in the pattern will be dominant in determining the subject score on that PC. This does not mean that weaker voxel weights should be discarded. "Although highly weighted regions may have a greater influence on the pattern score, regional values alone are not as predictive as whole pattern expression in performance measures" (Spetsieris et al. 2015). PCs are vectors, which do not have a direction. If controls have higher subject scores on a PC compared to patients, then the PC map is multiplied with −1 by convention. Subsequently, positive voxel weights are color-coded red in PC maps, and negative voxel weights are color-coded blue.

In order to interpret the pattern topography itself, it is useful to apply a threshold to the PC image, which somehow indicates which regions are most important. Several approaches have been applied in literature. For instance, all voxel values in the PC map can be transformed to Z-values. Next, only the highest and lowest voxel values in this "Z-map" are displayed (at a certain threshold, for instance, $Z > 1.96$ corresponding to $P < 0.05$). Regions that survive this threshold are interpreted as the most important regions of the pattern; these regions are likely most involved in the disease process.

An important issue arises when small sample sizes (20/20) are used, as is usually the case in neuroimaging studies. Pattern maps will likely be variable depending on

the specific sample of patients and controls. Thus, voxel weights in the pattern may fluctuate (this is also the case for univariate approaches). Identification of the areas most affected by the disease process becomes less reliable. This could be solved by collecting many different datasets of controls and patients and determining the disease-related pattern for each dataset. Imagine we have 1000 datasets of 20 controls and 20 patients. Then we can derive 1000 patterns. We can study the distribution of voxel weights across the different patterns. For each voxel, we can determine the mean voxel weight and its standard deviation. If a particular voxel has a similar weight in each pattern and does not fluctuate much across populations (i.e., it has a small standard deviation), then this voxel gives a reliable contribution to the pattern and is interpreted as being important in the disease. In contrast, if a voxel has a negative weight in some populations, but a positive voxel weight in others (i.e., the distribution straddles 0), then this voxel is probably unreliable. In summary, if we have such a distribution of voxel weights for each voxel, we can test which voxels in the pattern are reliable.

In reality, such a study is hampered by small numbers of subjects in most datasets and varying imaging protocols across centers which impedes pooling of datasets. To approximate this distribution per voxel, we apply a bootstrap estimation procedure. A bootstrap estimation procedure entails repeating the PCA several (~1000) times on randomly sampled data (with replacement) from the pool of patients and controls. In each iteration, the control and patient group contain the same number of images. This means that some subjects are represented more than once in some iterations, whereas others are completely omitted. We are thus creating multiple datasets from just one derivation set, which is similar to the LOOCV procedure. In contrast to the LOOCV, a bootstrap allows us to reuse subjects. For each iteration, each voxel will get a voxel weight. With multiple iterations, it is possible to analyze the distribution of weights of a voxel with a point estimate (average) of w and a standard deviation of s_w. Using this distribution, we can determine thresholds, which can subsequently be used to display only those voxels that we consider stable enough to be interpreted as part of the disease topography.

Again, one approach to this threshold is to make a Z-map using the bootstrap distribution. Each voxel will receive a z-value: $z = w/s_w$. "Sufficiently small variability of a voxel weight around its point estimate results in a Z-value of large magnitude, and indicates a reliable contribution to the covariance pattern" (Habeck et al. 2008). One can chose a certain z-threshold to show only stable voxels (e.g., $|z| > 1.96$ corresponding to a P-value of <0.05).

In previous studies, we chose to apply the confidence interval (CI) as a threshold. For each average voxel weight (w), we can determine a CI interval based on the distribution of the voxel weights from the bootstrap. This confidence interval (e.g., 90%) has an upper and a lower bound. For positive regions, we display only those voxels for which the lower bound of the confidence interval is larger than zero. For the negative regions, we display only those voxels for which the upper bound of the confidence interval is smaller than zero. Voxels for which the confidence interval straddles zero are not visualized.

The maps that follow from the bootstrap estimation procedure are only used for visualization. It is likely that those voxels that survive the chosen threshold after a bootstrap estimation procedure are truly related to the disease process and not coincidentally found in a given sample. Thus, if a pattern were to be identified in a completely new set of controls and patients, it is likely that those regions would be identified again.

In summary, several validation procedures are necessary when identifying a disease-related pattern with SSM PCA. First, it is important to calculate its expression in a new dataset to ascertain that the pattern can be generalized to new subjects. If a validation set is not available, a leave-one-out cross validation can be applied. This approach tests whether the pattern is a good predictor of the disease in a new subject. Ideally, subject scores are also correlated to some aspects of the disease, such as disease duration or severity of symptoms. Finally, the stability of the pattern is assessed with a bootstrap estimation procedure. Stable regions will probably be found again if the pattern was to be re-derived in a completely new population and may thus be interpreted as important regions in the pathophysiology of the disease.

4.3.8   Advantages of SSM PCA Over Univariate SPM Models

In reality, neither mass-univariate nor SSM PCA can provide an exact description of the pathophysiological mechanisms that underlie the disease and give rise to alterations in neuroimaging signals (Moeller and Habeck 2006). That said, SSM PCA has a few advantages over univariate approaches. The core issue in univariate approaches is that they assume that voxels are independent, which is not the case. The signal in neuroimaging data stems from the communication between neurons over synapses. The strength of neuroimaging lies in the detection of these interactions, to ultimately understand the network-level changes in certain conditions. Disregarding interactions between voxels means that most of the data of interest is in fact discarded (O'Toole et al. 2007). Multivariate approaches such as SSM PCA take into account the interactions (covariance) between voxels, and patterns that result from these analyses are more easily interpreted in the context of network-level changes.

The assumption of independent voxels in mass-univariate analyses also leads to a technical issue. In voxel-by-voxel comparisons, 10^4 to 10^5 voxels are compared between two groups, with an equal number of t tests. Performing multiple t tests leads to an inflation of the error rate. In such cases the α-level has to be corrected for multiple comparisons. This correction can be either too liberal (leading to type I errors) or too conservative (leading to type II errors), potentially "correcting away" true effects of interest in the data. Some solutions have been proposed (Genovese et al. 2002), but investigators are often willing to tolerate higher ($p > 0.05$) false-positive rates. For example, voxel clusters at uncorrected levels of $P < 0.001$ are often reported for standard F- or T-statistics (Moeller and Habeck 2006). In contrast, multivariate analyses such as SSM PCA have enhanced statistical power, as correction for multiple comparisons is not necessary.

In addition, it is generally accepted that PCA-based multivariate approaches have better sensitivity and replicability compared to univariate approaches (Habeck et al. 2008; Habeck et al. 2010). SSM PCA is especially relevant when making predictions in new cases. Habeck et al. compared SSM PCA pattern analysis to univariate approaches in the study of Alzheimer's disease (AD). In the univariate approach, AD patients and controls were compared using SPM with a t test. The area that gave the best group separation in the derivation sample was chosen as the region of interest (ROI). The ROI in this study was the right parietotemporal area. Signal values in this ROI were inspected in new AD subjects, which were subsequently classified as AD if the signal in that ROI surpassed a fixed threshold level obtained from the derivation sample. The multivariate approach entailed the identification of the AD-related pattern (ADRP) with SSM PCA. ADRP subject scores were calculated in scans of new AD patients, and these subjects were classified as AD if the subject score surpassed a fixed threshold level obtained from the derivation sample. Although both methods were able to distinguish between patients and controls in the derivation sample, the classification of new subjects was significantly better when the ADRP was used. Even when the most important area was omitted from the scans (i.e., the voxel values in the right parietotemporal area were set to zero in each subject), ADRP subject scores remained stable. According to the authors, "this demonstrates how multivariate analysis takes into account the interregional correlation structure in the data, and is thus not critically dependent on the inclusion of any particular brain region and can withstand dropping out even the most salient areas" (Habeck et al. 2008). This also supports the concept that pathological processes have a widely distributed effect on brain function in neurodegeneration.

Appendix: Effects of Normalization

In this example we demonstrate effects of ratio normalization versus log transformation and subtraction of the mean, as applied in the scaled subprofile model (SSM). In this example, we consider ¹⁸F-FDG uptake in two regions, A and B, in healthy controls and patients. Region B is affected by the disease. Metabolism in this region has changed compared to the control population with ΔB. In our example, region A is unaffected in both groups. For each subject, there is a scaling factor q which accounts for effects due to, for instance, the amount of radioactive label administered. The term q is a subject-specific scaling factor which we need to eliminate from the data.

Controls		Patients	
Region A	Region B	Region A	Region B
$q_1 A$	$q_1 B$	$q_2 A$	$q_2 (B + \Delta B)$

We introduce another term n, which indicates the relative size of region A.

$$\text{Mean brain uptake}: nA + (1-n)B \qquad\qquad 0 < n < 1$$

Often, ^{18}F-FDG uptake is altered in a few areas due to the disease. In such a case, A will contribute much more to the whole-brain average than B. Situations where B contributes only a small proportion to the whole-brain average include, for instance, early Parkinson's disease, or even its prodromal phases.

One can imagine situations where large parts of the brain are affected by the disease. For instance, in advanced Alzheimer's disease, we would expect hypometabolism in large parts of the cerebral cortex. In that case, B contributes overwhelmingly to the whole-brain average. As a result, whole-brain and average ^{18}F-FDG uptake will be lower in patients compared to controls.

As discussed in the main text, a normalization is needed. Two options are discussed. The first method is global mean normalization by proportional scaling, which is commonly applied and entails dividing each voxel value by the subject mean. The second method is the normalization procedure as applied in SSM PCA. In the SSM, data is first log-transformed and subsequently the subject mean is subtracted. In this example we will show that:

1. Scaling effects (q) are eliminated in both methods.
2. Both normalization techniques can introduce artifacts in (the unaffected) region A.

Global Mean Normalization

In global mean normalization, ^{18}F-FDG uptake in each voxel is divided by the mean uptake of the whole brain. For our example, the corrected values are shown in below figure
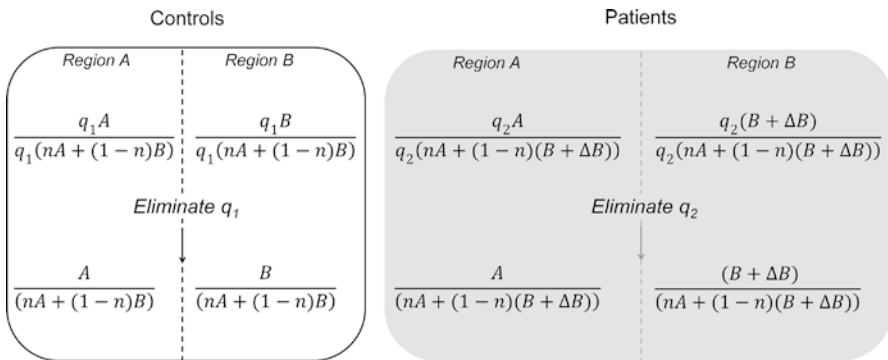
Although ^{18}F-FDG uptake in region A is the same in the control and patient population, the values in this area are different after global mean normalization.

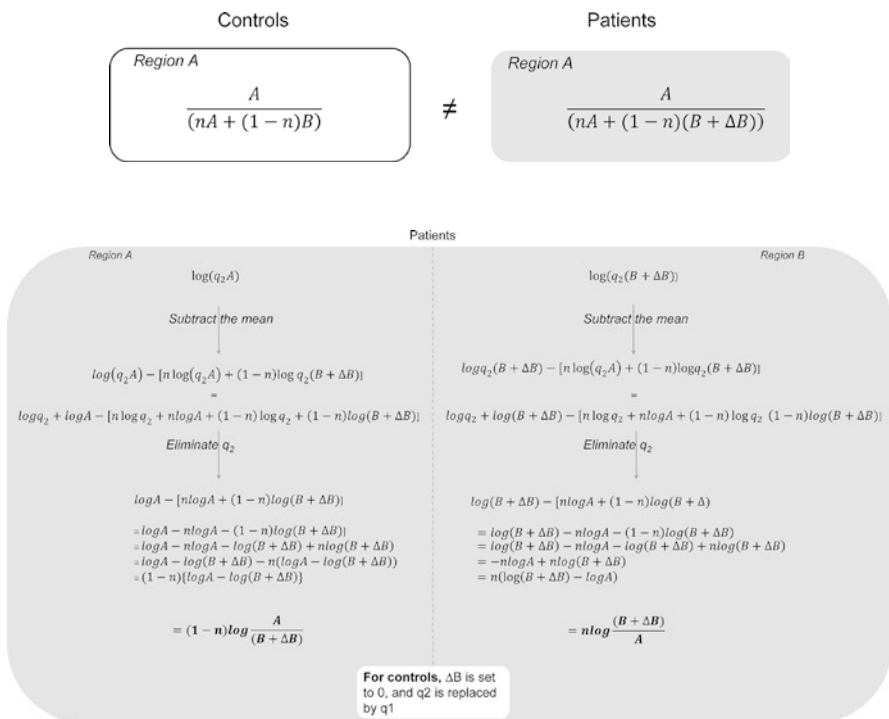Thus, the change in signal in region B due to pathology resulted in an altered signal in region A after global mean normalization (i.e., it produced an artifact).

Log Transformation and Demean in the SSM

In the SSM, the data are first log-transformed, and next we subtract the mean (see below figure). The fact that factor q can be eliminated in the SSM indicates that any multiplicative effect in the data can be removed, just like it can be eliminated in the global mean normalization. Thus, both methods are invariant to scaling effects (also see Spetsieris and Eidelberg (2011)).

Practical Examples

We modeled the formulas above in MATLAB, with two values for n (0.9 and 0.1) and variable values for ΔB. For A and B, we chose the same (realistic) fixed values.

Controls

Region A
$$\frac{A}{(nA + (1-n)B)}$$

$\neq$

Patients

Region A
$$\frac{A}{(nA + (1-n)(B + \Delta B))}$$

Patients

Region A
$$\log(q_2 A)$$

Subtract the mean

$$\log(q_2 A) - [n \log(q_2 A) + (1-n)\log q_2(B + \Delta B)]$$
$$=$$
$$\log q_2 + \log A - [n \log q_2 + n \log A + (1-n)\log q_2 + (1-n)\log(B + \Delta B)]$$

Eliminate q_2

$$\log A - [n \log A + (1-n)\log(B + \Delta B)]$$
$$= \log A - n \log A - (1-n)\log(B + \Delta B)]$$
$$= \log A - n \log A - \log(B + \Delta B) + n \log(B + \Delta B)$$
$$= \log A - \log(B + \Delta B) - n(\log A - \log(B + \Delta B))$$
$$= (1-n)\{\log A - \log(B + \Delta B)\}$$

$$= (1-n)\log \frac{A}{(B + \Delta B)}$$

Region B
$$\log(q_2(B + \Delta B))$$

Subtract the mean

$$\log q_2(B + \Delta B) - [n \log(q_2 A) + (1-n)\log q_2(B + \Delta B)]$$
$$=$$
$$\log q_2 + \log(B + \Delta B) - [n \log q_2 + n \log A + (1-n)\log q_2 \ (1-n)\log(B + \Delta B)]$$

Eliminate q_2

$$\log(B + \Delta B) - [n \log A + (1-n)\log(B + \Delta)]$$
$$= \log(B + \Delta B) - n \log A - (1-n)\log(B + \Delta B)$$
$$= \log(B + \Delta B) - n \log A - \log(B + \Delta B) + n \log(B + \Delta B)$$
$$= -n \log A + n \log(B + \Delta B)$$
$$= n(\log(B + \Delta B) - \log A)$$

$$= n \log \frac{(B + \Delta B)}{A}$$

For controls, ΔB is set to 0, and q2 is replaced by q1

Example 1: Changes in a few regions.

In this situation, ^{18}F-FDG changes are present in a few brain regions. In patients, most of the brain is unchanged, and thus A contributes most to the average (n is close to 1).

We plotted the values for A and B after global mean normalization ("A_{mean}" and "B_{mean}") and after SSM normalization ("A_{log}" and "B_{log}"). On the x-axis, we show the values for ΔB, ranging from -1000 (i.e., a decrease in B) to $+1000$ (i.e., an increase in B). Furthermore, we chose: $A = 1001$, $n = 0.9$ and $B = 1001$. The result is shown in Fig. 4.11a.

It is clear that there is an offset difference between the two methods. This is inherent to subtracting the mean versus dividing by the mean. When region B becomes hypometabolic, there is a slight (artificial) increase in region A. However, the changes in region A as a function of ΔB, even for extreme values of ΔB, are relatively small. The slope for the new values in A and B after each normalization procedure are almost equal.

Example 2: Changes in most of the brain.

In this situation, most of the cortex of the brain shows altered ^{18}F-FDG uptake in patients. Only a few brain regions have intact ^{18}F-FDG uptake (A), and these brain regions only contribute marginally to the whole-brain average. The altered brain areas (B) dominate the whole-brain metabolism, and ΔB is large. To simulate this situation, we repeated the example ($A = 1001$, $B = 1001$), but this time we chose $n = 0.1$. The result is shown in Fig. 4.11b. This example illustrates that both methods can cause an artifactual increase in A, when there is extreme hypometabolism in B.

To summarize, the grand mean normalization and the normalization in the SSM are equivalent methods. We illustrated that normalization to any mean is useful to eliminate subject-specific scaling factors in ^{18}F-FDG-PET data, but inherently can induce artificial increases and decreases. This is a known issue in any imaging study where absolute values are not available, be it univariate or multivariate. It is therefore important that patients and controls have similar values of average ^{18}F-FDG brain uptake (i.e., global metabolic rate (GMR)).

This issue has been addressed in several publications concerning the spatial covariance pattern that was identified in Parkinson's disease (Parkinson's disease-related pattern, PDRP). This pattern is characterized by relatively increased metabolism in subcortical structures (globus pallidus, putamen, thalamus, cerebellum, and pons), relatively increased metabolism in the sensorimotor cortex, and relatively decreased metabolism in the lateral frontal and parieto-occipital areas (Fig. 4.12). It has been posited that the PDRP reflects normalization artifacts due to GMR differences between controls and patients (Borghammer et al. 2008; Borghammer et al. 2009). Specifically, widespread cortical decreases, rather than subcortical increases,
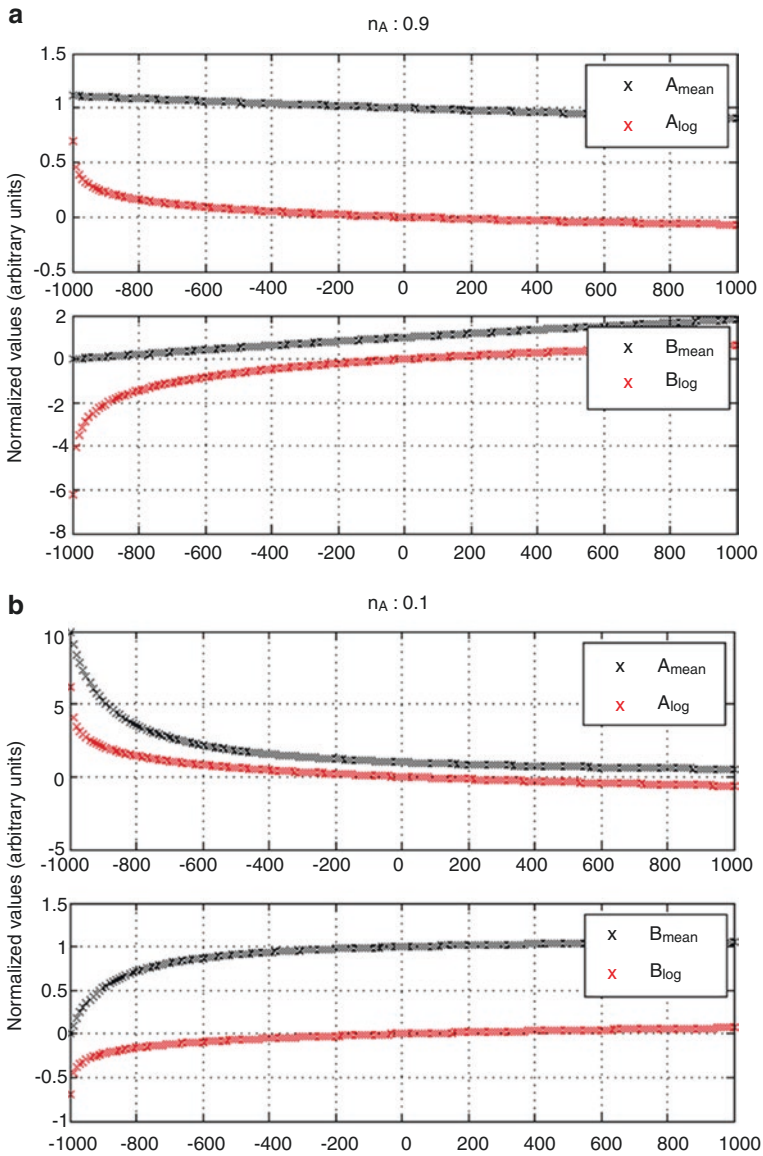
Fig. 4.11 The normalized values for region A are plotted for the global mean normalization method (A_{mean}: black) and for the SSM (A_{log}: red). The normalized values for region B are also plotted for both regions (B_{mean} and B_{log}). In a, the results of example 1 are shown ($n = 0.9$; $A = 1001$ and ΔB ranges from -1000 to 1000). In b, the results of example 2 are shown ($n = 0.1$; $A = 1001$ and ΔB ranges from -1000 to $+1000$)

PDRP
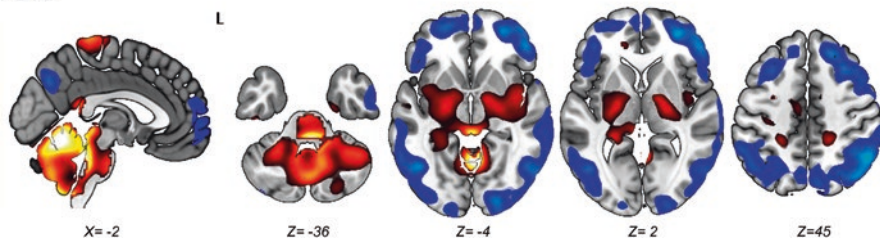
Fig. 4.12 The Parkinson's disease-related pattern (PDRP) identified in 17 controls and 19 PD subjects. Stable voxels are displayed, determined after a bootstrap resampling (90% confidence interval not straddling zero). Overlay on a T_1 MRI template. Positive voxel weights are color-coded red (relative hypermetabolism), and negative voxel weights are color-coded blue (relative hypometabolism). L = Left. Coordinates in the axial (Z) and sagittal (X) planes are in Montreal Neurological Institute (MNI) space

were suggested to be characteristic of the PD disease process (Borghammer et al. 2010). However, both theoretical and empirical evidence is available to support the contention that the PDRP topography holds true pathophysiological meaning and that the "red PDRP nodes" are central to PD pathophysiology.

Spetsieris et al. showed that GMR reductions in PD patients were not significant relative to healthy controls after 15 years of illness. GMR reductions also did not correlate with symptom duration (in contrast to PDRP scores) (Spetsieris and Eidelberg 2011). Ma et al. analyzed absolute ^{18}F-FDG uptake (with arterial blood sampling) in 24 patients with early-stage PD (Hoehn and Yahr I–II) and 24 controls. Both absolute (physiological units) and relative (after global mean ratio normalization) scan data was analyzed with a univariate model (SPM). A group contrast of relative count data revealed increased metabolism in the globus pallidus, ventral thalamus, dorsal pons/midbrain, and sensorimotor cortex, but cortical metabolic decreases were not found. There was no significant difference in mean whole-brain CMRglc between patients and controls. When absolute measures (physiological units without global mean normalization) were compared between groups in a similar univariate SPM analysis, no differences were found between controls and patients. This was attributed to the marked reduction in between-subject variability achieved with the normalization step. A similar analysis was also performed in repeat scans of PD patients. Globally normalized values for the "hypermetabolic regions" showed greater reproducibility than the corresponding absolute values (in physiological units). Thus, instead of introducing bias, the authors concluded that, when the global metabolic rate is carefully matched across groups, global normalization enhances the sensitivity of PET to detect meaningful regional differences. The SSM PCA disease-related pattern that was identified in the same data was similar to the SPM pattern but also included some additional regions (Ma et al. 2009).

A PDRP has also been identified by first normalizing the data to the cerebellum (non-log; every voxel divided by average cerebellar uptake), which was very similar to the original PDRP. In addition, the "red" and "blue" parts of the PDRP have also been used as separate vectors to calculate subject scores. Interestingly, both were

able to discriminate between controls and PD patients of a new dataset, in which the red pattern performed the best (Spetsieris and Eidelberg 2011). In addition, when the "red" and "blue" vectors were calculated separately in longitudinal FDG PET scans of de novo PD patients and controls (three scans per subject over a 48-month period), the rate of progression of the red regions was the greatest and significantly higher compared to controls. By contrast, the expression of the blue pattern did not differ from controls at any of the three time points (Ma et al. 2009). Moreover, if the "blue" areas in the PDRP define or cause the "red" areas in the PDRP, then the "red" areas should disappear when the PDRP is re-derived in a subspace that excludes the "blue" areas. This was not the case; a PDRP derived in the red voxel subspace was very similar to the "red" vector of the original PDRP, and subject scores for these two patterns were significantly correlated (Spetsieris and Eidelberg 2011). Finally, Dhawan et al. studied a group of healthy participants in whom global metabolic activity was experimentally decreased by sleep induction (with secobarbital). Participants were scanned with ¹⁸F-FDG PET while awake and during stage II sleep (monitored with EEG recordings). Sleep-induced reductions in global metabolic activity did not increase PDRP expression in these controls. In addition, an SSM PCA pattern comparing sleep and wake scans did not disclose any PDRP-like subcortical increases (Dhawan et al. 2012).

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723

Attwell D, Laughlin SB (2001) An energy budget for signaling in the grey matter of the brain. J Cereb Blood Flow Metab 21:1133–1145

Batipps M, Miyaoka M, Shinohara M, Sokoloff L, Kennedy C (1981) Comparative rates of local cerebral glucose utilization in the visual system of conscious albino and pigmented rats. Neurology 31:58–62

Borghammer P, Jonsdottir KY, Cumming P, Ostergaard K, Vang K, Ashkanian M, Vafaee M, Iversen P, Gjedde A (2008) Normalization in PET group comparison studies--the importance of a valid reference region. NeuroImage 40:529–540

Borghammer P, Cumming P, Aanerud J, Gjedde A (2009) Artefactual subcortical hyperperfusion in PET studies normalized to global mean: lessons from Parkinson's disease. NeuroImage 45:249–257

Borghammer P, Chakravarty M, Jonsdottir KY, Sato N, Matsuda H, Ito K, Arahata Y, Kato T, Gjedde A (2010) Cortical hypometabolism and hypoperfusion in Parkinson's disease is extensive: probably even at early disease stages. Brain Struct Funct 214:303–317

Bostan AC, Dum RP, Strick PL (2013) Cerebellar networks with the cerebral cortex and basal ganglia. Trends Cogn Sci 17:241–254

Buzsaki G, Kaila K, Raichle M (2007) Inhibition and brain work. Neuron 56:771–783

Buzsaki G, Logothetis N, Singer W (2013) Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. Neuron 80:751–764

Cherry SR, Sorenson JA, Phelps MA (2012) Physics in nuclear medicine. Elsevier Saunders, Philadelphia

Della Rosa PA, Cerami C, Gallivanone F, Prestia A, Caroli A, Castiglioni I, Gilardi MC, Frisoni G, Friston K, Ashburner J, Perani D, EADC-PET Consortium (2014) A standardized [18F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. Neuroinformatics 12:575–593

Dhawan V, Tang CC, Ma Y, Spetsieris P, Eidelberg D (2012) Abnormal network topographies and changes in global activity: absence of a causal relationship. NeuroImage 63:1827–1832

Eckert T, Barnes A, Dhawan V, Frucht S, Gordon MF, Feigin AS, Eidelberg D (2005) FDG PET in the differential diagnosis of parkinsonian disorders. NeuroImage 26:912–921

Eidelberg D (2009) Metabolic brain networks in neurodegenerative disorders: a functional imaging approach. Trends Neurosci 32:548–557

Eidelberg D, Moeller JR, Dhawan V, Spetsieris P, Takikawa S, Ishikawa T, Chaly T, Robeson W, Margouleff D, Przedborski S (1994) The metabolic topography of parkinsonism. J Cereb Blood Flow Metab 14:783–801

Fornito A, Zalesky A, Bullmore E (2016) Fundamentals of brain network analysis. Elsevier, Amsterdam

Fox PT, Raichle ME (1986) Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci U S A 83:1140–1144

Fox PT, Mintun MA, Raichle ME, Herscovitch P (1984) A noninvasive approach to quantitative functional brain mapping with H2 (15)O and positron emission tomography. J Cereb Blood Flow Metab 4:329–333

Fox PT, Raichle ME, Mintun MA, Dence C (1988) Nonoxidative glucose consumption during focal physiologic neural activity. Science 241:462–464

Friston KJ (2011) Functional and effective connectivity: a review. Brain Connect 1:13–36

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15:870–878

Habeck C, Moeller JR (2011) Intrinsic functional-connectivity networks for diagnosis: just beautiful pictures? Brain Connect 1:99–103

Habeck C, Foster NL, Perneczky R, Kurz A, Alexopoulos P, Koeppe RA, Drzezga A, Stern Y (2008) Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. NeuroImage 40:1503–1515

Habeck C, Stern Y, Alzheimer's Disease Neuroimaging Initiative (2010) Multivariate data analysis for neuroimaging data: overview and application to Alzheimer's disease. Cell Biochem Biophys 58:53–67

Heiss WD (2014) Cerebral Glucose Metabolism. In: Dierckx RAJO, Otte A, de Vries EFJ, van Waarde A, Luiten PGM (eds) PET and SPECT of neurobiological systems. Springer-Verlag, Berlin Heidelberg, p 85

Herholz K, Herscovitch P, Heiss W (2004) NeuroPET; PET in neuroscience and clinical neurology. Springer-Verlag, Berlin Heidelberg

Jueptner M, Weiller C (1995) Review: does measurement of regional cerebral blood flow reflect synaptic activity? Implications for PET and fMRI. NeuroImage 2:148–156

Juh R, Kim J, Moon D, Choe B, Suh T (2004) Different metabolic patterns analysis of parkinsonism on the 18F-FDG PET. Eur J Radiol 51:223–233

Kadekaro M, Crane AM, Sokoloff L (1985) Differential effects of electrical stimulation of sciatic nerve on metabolic activity in spinal cord and dorsal root ganglion in the rat. Proc Natl Acad Sci U S A 82:6010–6013

Karp JS, Surti S, Daube-Witherspoon ME, Muehllehner G (2008) Benefit of time-of-flight in PET: experimental and clinical results. J Nucl Med 49:462–470

Kogan RV, de Jong BA, Renken RJ, Meles SK, van Snick PJH, Golla S, Rijnsdorp S, Perani D, Leenders KL, Boellaard R, JPND-PETMETPAT Working Group (2019) Factors affecting the harmonization of disease-related metabolic brain pattern expression quantification in [(18) F]FDG-PET (PETMETPAT). Alzheimers Dement (Amst) 11:472–482

Lin AL, Fox PT, Yang Y, Lu H, Tan LH, Gao JH (2008) Evaluation of MRI models in the measurement of CMRO2 and its relationship with CBF. Magn Reson Med 60:380–389

Lin AL, Fox PT, Yang Y, Lu H, Tan LH, Gao JH (2009) Time-dependent correlation of cerebral blood flow with oxygen metabolism in activated human visual cortex as measured by fMRI. NeuroImage 44:16–22

Lin AL, Fox PT, Hardies J, Duong TQ, Gao JH (2010) Nonlinear coupling between cerebral blood flow, oxygen consumption, and ATP production in human visual cortex. Proc Natl Acad Sci U S A 107:8446–8451

Ma Y, Tang C, Spetsieris PG, Dhawan V, Eidelberg D (2007) Abnormal metabolic network activity in Parkinson's disease: test-retest reproducibility. J Cereb Blood Flow Metab 27:597–605

Ma Y, Tang C, Moeller JR, Eidelberg D (2009) Abnormal regional brain function in Parkinson's disease: truth or fiction? NeuroImage 45:260–266

Magistretti PJ, Allaman I (2015) A cellular perspective on brain energy metabolism and functional imaging. Neuron 86:883–901

Meles SK, Kok JG, De Jong BM, Renken RJ, de Vries JJ, Spikman JM, Ziengs AL, Willemsen ATM, van der Horn HJ, Leenders KL, Kremer HPH (2018a) The cerebral metabolic topography of spinocerebellar ataxia type 3. Neuroimage Clin 19:90–97

Meles SK, Renken RJ, Janzen A, Vadasz D, Pagani M, Arnaldi D, Morbelli S, Nobili F, Mayer G, Leenders KL, Oertel WHO (2018b) The metabolic pattern of idiopathic REM sleep behavior disorder reflects early-stage Parkinson's disease. J Nucl Med 59(9):1437–1444

Moeller JR, Habeck CG (2006) Reciprocal benefits of mass-univariate and multivariate modeling in brain mapping: applications to event-related functional MRI, H(2) (15)O-, and FDG-PET. Int J Biomed Imaging 2006:79862

Moeller JR, Strother SC (1991) A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. J Cereb Blood Flow Metab 11:A121–A135

Moeller JR, Strother SC, Sidtis JJ, Rottenberg DA (1987) Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. J Cereb Blood Flow Metab 7:649–658

Niethammer M, Eidelberg D (2012) Metabolic brain networks in translational neurology: concepts and applications. Ann Neurol 72:635–647

O'Toole AJ, Jiang F, Abdi H, Penard N, Dunlop JP, Parent MA (2007) Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. J Cogn Neurosci 19:1735–1752

Phelps ME, Hoffman EJ, Huang SC, Ter-Pogossian MM (1975) Effect of positron range on spatial resolution. J Nucl Med 16:649–652

Raichle ME, Larson KB, Phelps ME, Grubb RL Jr, Welch MJ, Ter-Pogossian MM (1975) In vivo measurement of brain glucose transport and metabolism employing glucose- -11C. Am J Phys 228:1936–1948

Reivich M, Kuhl D, Wolf A, Greenberg J, Phelps M, Ido T, Casella V, Fowler J, Hoffman E, Alavi A, Som P, Sokoloff L (1979) The [18F]fluorodeoxyglucose method for the measurement of local cerebral glucose utilization in man. Circ Res 44:127–137

Rodriguez-Oroz MC, Jahanshahi M, Krack P, Litvan I, Macias R, Bezard E, Obeso JA (2009) Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. Lancet Neurol 8:1128–1139

Sacks W (1957) Cerebral metabolism of isotopic glucose in normal human subjects. J Appl Physiol 10:37–44

Sibson NR, Dhankhar A, Mason GF, Rothman DL, Behar KL, Shulman RG (1998) Stoichiometric coupling of brain glucose metabolism and glutamatergic neuronal activity. Proc Natl Acad Sci U S A 95:316–321

Sokoloff L (1993) Sites and mechanisms of function-related changes in energy metabolism in the nervous system. Dev Neurosci 15:194–206

Sokoloff L, Reivich M, Kennedy C, Des Rosiers MH, Patlak CS, Pettigrew KD, Sakurada O, Shinohara M (1977) The [^{14}C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. J Neurochem 28:897–916

Spetsieris PG, Eidelberg D (2011) Scaled subprofile modeling of resting state imaging data in Parkinson's disease: methodological issues. NeuroImage 54:2899–2914

Spetsieris PG, Ko JH, Tang CC, Nazem A, Sako W, Peng S, Ma Y, Dhawan V, Eidelberg D (2015) Metabolic resting-state brain networks in health and disease. Proc Natl Acad Sci U S A 112:2563–2568

Surti S, Karp JS (2016) Advances in time-of-flight PET. Phys Med 32:12–22

Teune LK, Bartels AL, de Jong BM, Willemsen AT, Eshuis SA, de Vries JJ, van Oostrom JC, Leenders KL (2010) Typical cerebral metabolic patterns in neurodegenerative brain diseases. Mov Disord 25:2395–2404

Teune LK, Renken RJ, Mudali D, De Jong BM, Dierckx RA, Roerdink JB, Leenders KL (2013) Validation of parkinsonian disease-related metabolic brain patterns. Mov Disord 28:547–551

Teune LK, Renken RJ, de Jong BM, Willemsen AT, van Osch MJ, Roerdink JB, Dierckx RA, Leenders KL (2014a) Parkinson's disease-related perfusion and glucose metabolic brain patterns identified with PCASL-MRI and FDG-PET imaging. Neuroimage Clin 5:240–244

Teune LK, Strijkert F, Renken RJ, Izaks GJ, de Vries JJ, Segbers M, Roerdink JB, Dierckx RA, Leenders KL (2014b) The Alzheimer's disease-related glucose metabolic brain pattern. Curr Alzheimer Res 11:725–732

Vaishnavi SN, Vlassenko AG, Rundle MM, Snyder AZ, Mintun MA, Raichle ME (2010) Regional aerobic glycolysis in the human brain. Proc Natl Acad Sci U S A 107:17757–17762

Varrone A, Asenbaum S, Vander Borght T, Booij J, Nobili F, Nagren K, Darcourt J, Kapucu OL, Tatsch K, Bartenstein P, Van Laere K, European Association of Nuclear Medicine Neuroimaging Committee (2009) EANM procedure guidelines for PET brain imaging using [18F]FDG, version 2. Eur J Nucl Med Mol Imaging 36:2103–2110

Wu P, Wang J, Peng S, Ma Y, Zhang H, Guan Y, Zuo C (2013) Metabolic brain network in the Chinese patients with Parkinson's disease based on 18F-FDG PET imaging. Parkinsonism Relat Disord 19:622–627

Wu P, Yu H, Peng S, Dauvilliers Y, Wang J, Ge J, Zhang H, Eidelberg D, Ma Y, Zuo C (2014) Consistent abnormalities in metabolic network activity in idiopathic rapid eye movement sleep behaviour disorder. Brain 137:3122–3128

Yong SW, Yoon JK, An YS, Lee PH (2007) A comparison of cerebral glucose metabolism in Parkinson's disease, Parkinson's disease dementia and dementia with Lewy bodies. Eur J Neurol 14:1357–1362