Estimating functional brain maturity in very and extremely preterm neonates using automated analysis of the electroencephalogram

J.M. O'Toole, G.B. Boylan, S. Vanhatalo, N.J. Stevenson

Objective: To develop an automated estimate of EEG maturational age (EMA) for preterm neonates.

Methods: The EMA estimator was based on the analysis of hourly epochs of EEG from 49 neonates with gestational age (GA) ranging from 23 to 32 weeks. Neonates had appropriate EEG for GA based on visual interpretation of the EEG. The EMA estimator used a linear combination (support vector regression) of a subset of 41 features based on amplitude, temporal and spatial characteristics of EEG segments. Estimator performance was measured with the mean square error (MSE), standard deviation of the estimate (SD) and the percentage error (SE) between the known GA and estimated EMA.

Results: The EMA estimator provided an unbiased estimate of EMA with a MSE of 82 days (SD = 9.1 days; SE = 4.8%) which was significantly lower than a nominal reading (the mean GA in the dataset; MSE of 267 days, SD of 16.3 days, SE = 8.4%; p < 0.001). The EMA estimator with the lowest MSE used amplitude, spatial and temporal EEG characteristics.

Conclusions: The proposed automated EMA estimator provides an accurate estimate of EMA in early preterm neonates.

Significance: Automated analysis of the EEG provides a widely accessible, noninvasive and continuous assessment of functional brain maturity.

1. Introduction

Every year, over two million babies are born very or extremely premature, less than 32 weeks gestational age (GA), and will require admission to a neonatal intensive care unit (NICU) (Blencowe et al., 2012). Neurological complications from prematurity can result in a 10–25 fold increase in annual healthcare costs (Kancherla et al., 2012). While recent progress in
2. Method

2.1. Database

2.1.1. Subjects

A database of EEG recordings from 49 preterm neonates with appropriate EEG for GA was used to develop the automated EMA estimator. Neonates with a range of GA from 23 weeks plus 3 days to 32 weeks plus 0 days (164–224 days) were included in the database. The distribution of GA in the database is shown in Fig. 1(A); the mean GA was 28.6 weeks (198 days) with a standard deviation of 16 days. Neonates were enrolled for EEG monitoring from the NICU of the Cork University Maternity Hospital, Ireland from January 2009 to October 2011. Approval for the study was obtained from the Clinical Research Ethics Committees of the Cork Teaching hospitals, Ireland. Written, informed consent was received from at least one parent of each neonate included in the study.

2.1.2. EEG recording

Multi-channel, conventional video-EEG recording was commenced on enrolment (within 72 h of birth) and continued for up to 3 days. A Nicolet One EEG machine (Natus Medical Inc., Pleasanton, CA, USA) was used to acquire the EEG. An array of 10 scalp electrodes were placed according to the International 10–20 system of electrode placement modified for neonates: frontal (F3, F4), central (C3, C4, Cz), temporal (T3, T4), occipital (O1, O2), and a reference. A bipolar montage of 8 channels was used in this study: C4–O2, C3–O1, C4–T4, C3–T3, C4–C2, Cz–C3, F4–C4, F3–C3. Electrode to scalp impedance was maintained below 5 kΩ when possible. EEGs were recorded with a sampling frequency of 256 Hz. After each EEG was recorded, all identifiable patient information was removed from the recording and the EEG was stored with a unique study number.

2.1.3. EEG review

The EEGs were examined by an experienced neonatal neurophysiologist (GBB) and were included if the EEG was judged to be appropriate for MA (André et al., 2010; Aminoff, 2012) and no clear abnormalities were present on the EEG (Watanabe et al., 2014; Shany et al., 2014). Such analysis is, however, challenged by several caveats: (1) it is qualitative, (2) the required expertise and access to facilities are limited and (3) assessment is rarely performed in a spatial context.

A device that provides a computational means of tracking EEG brain maturation in preterm neonates, allowing comparison of the recorded and expected EEG maturation, would be a useful tool for clinicians in the NICU (Scher, 1997). Quantitative analyses have suggested a wide variety of signal properties, estimated from automated segmentations of the EEG and measurements of spectral power, amplitude and connectivity, that correlate with MA (Holthausen et al., 2000; Niemarkt et al., 2011; O’Reilly et al., 2012; Koolen et al., 2014; Meijer et al., 2014; Murphy et al., 2015; Saji et al., 2015; Schumacher et al., 2015). These analyses provide a candidate feature set for use in an automated EMA estimator in the very and extremely premature neonate.

In the present work, we developed an “EMA estimator” for preterm neonates, with appropriate EEG for GA, based on automated analysis of the EEG. We used a combination of well studied features that have been shown to correlate with MA extracted from different segments of the EEG recording. The performance of our EMA estimator was then evaluated on a relatively large database of preterm neonates to provide proof of concept evidence for the design of a novel automated EEG maturity index.
activities provide hallmarks for the maturation of the mechanisms that generate cortical activity (Vanhatalo et al., 2005; Tolonen et al., 2007; Myers et al., 2012). In addition to these state changes at multi-second scales, preterm babies are also known to exhibit fluctuation in vigilance or brain states at the range of tens of minutes. These states alternate between high and low EEG activity periods (indirectly representing active and quiet sleeps, or REM and non-REM, respectively), and are periodically interrupted by a short awake period. Since the accumulated duration of SATs is significantly higher in the active state, it is reasonable to assess the EEG with respect to these EEG activity states (Palmu et al., 2013). We, therefore, segmented the EEG into low and high activity states, state 1 and 2, respectively, using a SAT-based measure of cortical activity (Niemarkt et al., 2010; Palmu et al., 2013; Stevenson et al., 2014a). EEG features were extracted from the entire EEG recording and different activity states and then combined using...
support vector regression (SVR) to generate an EMA estimate, see Fig. 3.

2.2.1. Pre-processing and segmentation

The EEG signals were initially filtered with a Butterworth, 1st order, high-pass filter with a cut-off frequency of 0.5 Hz and then down-sampled, with anti-aliasing low pass filtering, to 64 Hz.

Periods of SAT and inter-SAT were segmented on each channel using the SAT detection algorithm of Palmu et al. (2010). We used a slight modification to the post-processing stage. In our case, we use a threshold of 5 μV^2, rather than 1.5 μV^2, in order to account for differences in the frequency response of the nonlinear energy operator applied to the EEG at a reduced sampling frequency. The output of this algorithm was a binary time-series where 1 denotes the presence of SAT and 0 the presence of inter-SAT and the minimum SAT duration was one second.

This automated annotation was utilised in two further pre-processing stages: EEG activity state segmentation and artefact detection.

The SAT annotations were used to determine EEG activity states via the SAT%, defined as,

$$ SAT\% (t) = \frac{1}{T} \int_{t-1/2}^{t+1/2} SAT(t_0) dt_0 $$

where \( t \) is time in seconds, \( T = 300 \text{ s} (5 \text{ min}) \) and \( SAT(t) \) is the binary SAT annotation, see Fig. 4(D). This signal was then de-trended using a linear function and then segmented using an adaptive threshold. The threshold was chosen to ensure that the accumulated duration of each EEG activity state was approximately equal. The assumption of such a threshold is that changes in EEG activity states were apparent in every recording and the accumulated duration of each EEG state was approximately equal. A final stage was implemented to ensure a minimum EEG activity state duration of 5 min.

Two forms of artefact were present in the EEG recordings of our dataset: high amplitude artefacts and low amplitude artefacts. Simple methods of detecting each artefact were implemented. High amplitude artefacts are typically caused by patient movement or disruption to the recording by clinical staff. In order to detect large amplitude artefacts the SAT periods were investigated. Discrete SAT events where the maximum absolute EEG value exceeded 500 μV on at least 2 out of 8 channels were assumed to be artefactual, see Fig. 4(B). Any detection of high amplitude artefact resulted in all 8 channels of EEG being eliminated from subsequent analysis. Low amplitude artefacts are typically caused by bridging between electrodes resulting in low activity on distinct channels of a bipolar montage (see Fig. 4A). Low amplitude was defined as an order of magnitude less (10^{-1}) than the median absolute amplitude of the entire EEG recording across all EEG channels.

Any detection of low amplitude artefact resulted in the detected EEG channels being ignored in subsequent analysis.

2.2.2. Feature extraction

Several features, expected to correlate with EMA, were extracted from the EEG. These measures can be generalised into three classes: amplitude, spatial organisation and temporal organisation.

Amplitude features represent the raw EEG voltage. We used two types of amplitude estimate: the signal envelope based on the analytic associate of a signal and the range EEG (rEEG) (O’Reilly et al., 2012). The signal envelope is \(|z(t)| = |x(t)| + jH(x(t))|\) where H is the Hilbert transform, j denotes a complex number and \(x(t)\) is the EEG signal. The rEEG is defined as maximum EEG value minus the minimum EEG value within a non-overlapping 2 s window (O’Reilly et al., 2012).

The amplitude was further measured across different frequency bands using the power spectral density. A periodogram with no smoothing was used to estimate the power spectral density. The frequency bands used in analysis were 0.5–3, 3–8, 8–15, and 15–30 Hz. These were modified from traditional band definitions to better suit the distribution of EEG energy with respect to frequency in preterm neonates (Tokariev et al., 2012). Relative band measures were defined by dividing the power in a specific frequency band by the total power in all frequency bands of interest.

$$ R_B = \frac{\int_B |X(f)|^2 df}{\int_0^\infty |X(f)|^2 df} $$

where \( R_B \) is the relative spectral power in band \( B \), \( B \) is the range of the band of interest, \( W \) ranges on [0.5, 30] Hz and \( X(f) \) is the Fourier transform of the EEG signal. The implementation of the SAT detector means that amplitude features were estimated on data from the entire EEG recording and on each SAT. When estimating spectral band power on a SAT by SAT basis, only SATs with a minimum duration of two seconds were used to ensure a minimum frequency resolution of 0.5 Hz.

Measures of the temporal organisation of the EEG were based on the SAT detection output. The algorithm was run independently on each channel of the EEG. The resultant annotation was summarised with measures such as SAT duration, inter-SAT interval (also referred to as inter-burst interval), and number of SATs per hour.

Measures of the spatial organisation of the EEG assessed the inter-hemispheric synchrony between EEG channels. We used an implementation of the activation symmetry index (Räsänen et al., 2013) and Pearson’s correlation between channel envelopes (\(|z(t)|\)). These measures were applied to 3 channel pairs (F3–C3/F4–C4, C4–T4/C3–T3, and C4–O2/C3–O1).

All measures were estimated on each channel, where applicable, and summarised over time or SAT segments. These values were then summarised across EEG channels using the median. In order to account for possible rapid postnatal adaption we also used the time after birth of the EEG recording as a feature. A summary of the features used are shown in Table 1. These features were summarised with a variety of statistics such as percentiles or the root mean square. Amplitude measures were estimated on all available data (12 features) or on SAT and inter-SAT periods (18 features). All temporal features were estimated directly from the automated SAT annotation (9 features). Only two spatial features were used. These 41 features were estimated on three different segmentations: (1) the entire EEG recording, (2) S1 (low activity or quiet sleep) or (3) S2 (high activity or active sleep); resulting in a total of 124 features (including postnatal age of EEG recording).

We assessed several combinations of feature class extracted from different data segments, selected \textit{a priori}, in order to assess
the efficacy of feature class and segmentation method on EMA performance.

2.2.3. EMA estimation using SVR

The features were combined using SVR to form an estimate of the EMA (Smola and Schölkopf, 2004; see Appendix A for more details). In order to implement the SVR, the support vectors, weights and bias term must be determined. In SVR, these values are found by minimising a cost function consisting of measures of flatness and empirical error calculated on training data. In training, the SVR parameters $C$ and $\varepsilon$ can also be optimally selected; due to limitations on the size of our dataset we used the default settings of $C = 1$ and $\varepsilon = 0.1$ found in the LIBSVM package for SVR (Chang and Lin, 2011).

2.3. Training and testing

We used cross-validation to assess the accuracy of the EMA estimate. The advantage of such an approach is that the EMA estimator is not trained on testing data. A leave-one subject out (LOSO) cross-validation was used. In this case, all data except that of a single neonate was included in the training set. The trained EMA estimator was then applied to the left out neonate to assess its performance. This was repeated until all neonates had been left-out of the training data. The performance of the EMA estimator was defined using several metrics: bias, mean square error (MSE), correlation coefficient ($\rho_{EMA}$), standard deviation of error in days (SD), and the standard deviation of the percentage error (SE) between the known GA and the estimated EMA.

The MSE was defined as,

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$  \hspace{1cm} (3)

where $\hat{y}_i$ is the estimated EMA, $y_i$ is the GA and $n$ is the number of observations ($n = 147$ epochs). The SD was defined as,

$$\text{SD} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (e_i - \bar{e})^2}$$  \hspace{1cm} (4)

where $e_i = \hat{y}_i - y_i$ and the SE was defined as,

$$\text{SE} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{e_i - \bar{e}}{y_i} \right)^2} \times 100$$  \hspace{1cm} (5)

The bias of the estimate was defined as the mean of $e_i$, that is $\bar{e}$.

Features were initially transformed with a Box–Cox transformation and then normalised by converting to z-scores (Box and Cox, 1964). The former is not a requirement for implementing a SVR but was performed to assist preliminary visualisation. The coefficients of each transformation were estimated in the training phase but was performed to assist preliminary visualisation. The coefficients of each transformation were estimated in the training phase and then applied to the test data. The efficacy of individual features for the estimation of GA was also estimated using LOSO cross-validation (EMA estimation via linear SVR). The correlation between individual features and the GA was also estimated using Spearman’s correlation coefficient ($\rho_{CA}$), see Fig. 5.

A feature selection method was not implemented in the development of the EMA estimator. Rather, we assessed several pre-selected combinations of features based on their class and segmentations in order to assess the viability of these factors on the overall EMA estimator.
Comparisons between different EMA estimates were performed using a paired t-test. Repeated measures in each neonate were reduced to a single summary measure per neonate using an average so that \( n = 49 \). The intra-patient variability was estimated using the standard deviation between three EMA estimates per neonate which were then summarised across neonates with a median and inter-quartile range.

### 3. Results

The artefact detection system eliminated high amplitude portions of EEG in 58 out of 147 EEG epochs. The median amount of data eliminated in these 58 epochs was 0.67% (IQR: 0.35–1.28). Channel shorting resulted in the removal of EEG channels in 22 out of 147 EEG epochs; no more than 2 channels were removed in a single EEG epoch.

Features with the lowest MSE between the estimated and actual EMA are shown in Table 2 (only one summary statistic and segmentation is included for each feature). The 5th percentile of the relative beta power calculated across all SATs in the EEG recording had the lowest MSE. An EMA estimator based this feature had a SD = 11.05 days (SE = 5.79%). The median absolute error of this estimator was 2.69% (IQR: 1.17–4.91%), which corresponds to a correlation of 0.833 (95% CI: 0.775–0.876; Pearson's correlation coefficient), see Fig. 5.

The performance of the EMA estimator based on a combination of features was significantly higher than the performance of a single feature (SD = 9.08 vs. SD = 11.05: \( p = 0.019 \), \( n = 49 \); paired t-test) and the mean GA in the dataset which provides a lower limit of possible performance (SD = 9.08 vs. SD = 16.30: \( p < 0.001 \), \( n = 49 \); paired t-test). The performance of the best EMA estimator features was significantly higher than an estimator based on only spatial or temporal features (\( p < 0.001 \) and \( p < 0.001 \), respectively, \( n = 49 \); paired t-test), but was not significantly higher than an estimator based on only amplitude features (\( p = 0.292 \), \( n = 49 \); paired t-test).

The median intra-patient variability was 4.42 days (IQR: 2.79–6.40). Re-evaluating the performance of the estimator by using an EMA averaged across the three time points per neonate reduced the error to SD = 7.85 days (SE = 4.22%) and resulted in a correlation 0.889 (95% CI: 0.811–0.936) between known GA and estimated EMA.

The performance of the EMA estimator with respect to weekly classification intervals is shown in Table 4. The visual interpretation of the EEG, the gold standard of EEG assessment of MA, has a reported accuracy of approximately SD = 12.42 days (SE = 5.35%) (Parmelee et al., 1968). EMA estimates deviated by more than two weeks from the GA in all three epochs of two neonates (neonates 8 and 32), two epochs from three neonates (neonates 4, 28 and 38) and one epoch in six neonates (neonates 3, 5, 6, 24, 29 and 35). Visual interpretation of the EEG in the two outliers (neonates 8 and 32) showed no obvious characteristics to

---

### Table 1

Features used in analysis. N/A implies that summarising is inherent in the calculation of the feature, SAT is spontaneous activity transient and RMS is root mean square. The spectral features estimated on the entire EEG segment produced a single value; however, when estimated on SATs the 5th, 50th and 95th percentiles were generated.

<table>
<thead>
<tr>
<th>Feature class</th>
<th>Feature</th>
<th>Summary statistics</th>
<th>Segments</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude</td>
<td>Envelope</td>
<td>5th, 50th, 95th</td>
<td>EEG, SAT, inter-SAT</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>rEEG</td>
<td>5th, 50th, 95th</td>
<td>EEG</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Total power (0.5–30 Hz)</td>
<td>N/A</td>
<td>EEG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Relative delta power (0.5–3 Hz)</td>
<td>N/A, 5th, 50th, 95th</td>
<td>EEG, SAT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Relative theta power (3–8 Hz)</td>
<td>N/A, 5th, 50th, 95th</td>
<td>EEG, SAT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Relative alpha power (8–15 Hz)</td>
<td>N/A, 5th, 50th, 95th</td>
<td>EEG, SAT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Relative beta power (15–30 Hz)</td>
<td>N/A, 5th, 50th, 95th</td>
<td>EEG, SAT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Temporal theta power (3–8 Hz)</td>
<td>N/A</td>
<td>EEG</td>
<td>1</td>
</tr>
<tr>
<td>Temporal</td>
<td>Number of SATs per hour</td>
<td>N/A</td>
<td>EEG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>SAT duration</td>
<td>5th, 50th, 95th, RMS</td>
<td>EEG</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Inter-SAT duration</td>
<td>5th, 50th, 95th, RMS</td>
<td>EEG</td>
<td>4</td>
</tr>
<tr>
<td>Spatial</td>
<td>Correlation of the envelope</td>
<td>N/A</td>
<td>EEG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Activation symmetry index</td>
<td>N/A</td>
<td>EEG</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>Postnatal age of EEG recording</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
</tr>
</tbody>
</table>

---

**Fig. 5.** Scatter plots of the relative beta power (RBP) with respect to gestational age. (A) The raw RBP value. (B) RBP after Box–Cox transformation using cross-validation. (C) RBP after Box–Cox transformation and support vector regression using cross-validation – the estimate of EEG maturational age (EMA).
Table 2
Estimation performance for individual features of EMA estimation. ES refers to the EEG activity state segmentation: none, S1 or S2. SAT refers to the SAT segmentation: yes or no. \( p_{\text{EMA}} \) is Pearson’s correlation between the EMA estimate and known GA using cross-validation and \( p_{\text{GA}} \) is Spearman’s correlation between the feature and the known GA.

<table>
<thead>
<tr>
<th>Feature Class</th>
<th>Feature</th>
<th>Percentile</th>
<th>ES</th>
<th>SAT</th>
<th>MSE (days(^2))</th>
<th>SD (days)</th>
<th>( p_{\text{EMA}} )</th>
<th>( p_{\text{GA}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative beta power</td>
<td>A</td>
<td>5th</td>
<td>None</td>
<td>Yes</td>
<td>122.93</td>
<td>11.05</td>
<td>0.739</td>
<td>0.721</td>
</tr>
<tr>
<td>Inter-SAT envelope</td>
<td>A</td>
<td>5th</td>
<td>S2</td>
<td>Yes</td>
<td>128.95</td>
<td>11.38</td>
<td>0.720</td>
<td>0.747</td>
</tr>
<tr>
<td>rEEG</td>
<td>A</td>
<td>5th</td>
<td>None</td>
<td>Yes</td>
<td>175.40</td>
<td>13.26</td>
<td>0.596</td>
<td>0.628</td>
</tr>
<tr>
<td>Correlation of the envelope</td>
<td>S</td>
<td>50th</td>
<td>S1</td>
<td>No</td>
<td>186.04</td>
<td>13.82</td>
<td>0.558</td>
<td>0.589</td>
</tr>
<tr>
<td>Envelope</td>
<td>A</td>
<td>5th</td>
<td>None</td>
<td>No</td>
<td>197.54</td>
<td>14.09</td>
<td>0.510</td>
<td>0.585</td>
</tr>
<tr>
<td>SATs per hour</td>
<td>T</td>
<td>N/A</td>
<td>None</td>
<td>Yes</td>
<td>210.06</td>
<td>14.41</td>
<td>0.477</td>
<td>0.506</td>
</tr>
<tr>
<td>Inter-SAT duration</td>
<td>T</td>
<td>95th</td>
<td>S1</td>
<td>Yes</td>
<td>229.35</td>
<td>15.12</td>
<td>0.397</td>
<td>-0.493</td>
</tr>
</tbody>
</table>

* Denotes a significantly increased mean square error (MSE) between a feature and the highest correlating feature (\( p < 0.05 \); paired t-test).

Table 3
Leave one-subject out cross-validation performance of the EMA estimates using several combinations of feature class. ES refers to the EEG state segmentation: none, S1 or S2. SAT refers to the SAT segmentation: yes or no.

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Feature number</th>
<th>ES</th>
<th>SAT</th>
<th>MSE (days(^2))</th>
<th>( \rho_{\text{TEST}} )</th>
<th>SD (days)</th>
<th>SE (%)</th>
<th>Bias (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13</td>
<td>S1</td>
<td>No</td>
<td>88.74</td>
<td>0.817</td>
<td>9.44</td>
<td>4.94</td>
<td>0.38</td>
</tr>
<tr>
<td>S’</td>
<td>3</td>
<td>S1</td>
<td>No</td>
<td>203.27</td>
<td>0.513</td>
<td>14.16</td>
<td>7.63</td>
<td>2.05</td>
</tr>
<tr>
<td>T’</td>
<td>15</td>
<td>S1</td>
<td>Yes</td>
<td>197.91</td>
<td>0.485</td>
<td>14.44</td>
<td>7.63</td>
<td>0.98</td>
</tr>
<tr>
<td>A, S</td>
<td>15</td>
<td>S1</td>
<td>No</td>
<td>83.52</td>
<td>0.830</td>
<td>9.14</td>
<td>4.72</td>
<td>0.70</td>
</tr>
<tr>
<td>A, T</td>
<td>22</td>
<td>S1</td>
<td>No</td>
<td>83.9</td>
<td>0.828</td>
<td>9.19</td>
<td>4.84</td>
<td>-0.05</td>
</tr>
<tr>
<td>A, T, S</td>
<td>24</td>
<td>S1</td>
<td>No</td>
<td>82.05</td>
<td>0.833</td>
<td>9.08</td>
<td>4.84</td>
<td>0.40</td>
</tr>
</tbody>
</table>

* Denotes a significant increase in the mean square error compared to the EMA estimate based on only amplitude (\( p < 0.05 \); paired t-test). All values of \( \rho_{\text{TEST}} \) are significantly different from 0 (\( p < 0.001 \)).

Fig. 6. Cross-validation output of the best performing EEG maturational age (EMA) estimate compared to the gestational age (GA). The black solid line denotes the output of an ideal EMA estimate and the broken red lines define increasing weekly deviations from the GA.

Table 4
The accuracy of the best performing EEG maturational age (EMA) estimator in terms of weekly intervals. Estimates are based on 147 h long EEG recordings taken from 49 neonates. The lower limit estimate is the mean of the GA in the dataset. To benchmark our findings with the literature, the visual interpretation (VI) of the EEG was taken from Parmelee et al. (1968); this was, however, based on EEG recordings from preterm neonates with a different age range.

<table>
<thead>
<tr>
<th>Estimate within</th>
<th>Best estimate (automated)</th>
<th>Lower limit (mean GA)</th>
<th>VI of EEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 week</td>
<td>89/147 (61%)</td>
<td>45/147 (31%)</td>
<td>11/20 (55%)</td>
</tr>
<tr>
<td>2 weeks</td>
<td>128/147 (87%)</td>
<td>78/147 (53%)</td>
<td>18/20 (90%)</td>
</tr>
</tbody>
</table>

4. Discussion

We have shown that automated EEG analysis yields an index of brain maturity with an accuracy that parallels, even outperforms, the conventional visual interpretation of the highly skilled expert (Parmelee et al., 1968). It is comparable in accuracy to anatomical measurement at birth (Hunter, 2009); we assume that the accuracy associated with ultrasound measurement of anatomical landmarks, a SE of 4%, will translate to physical measurements ex-utero. The best EMA estimate was obtained with linear SVR combining temporal, spatial and energy based features extracted from a low-activity segment of EEG. Our findings are compatible with previous studies that have reported changes in various EEG features with maturation; however, this is the first work to combine such features to directly compute maturation in preterm neonates. The combination of quantitative features in a Bayesian framework was recently used by Jakaite et al. (2012) for the estimation of conceptional age in term neonates. Their study focused on a cohort of term neonates ranging from 41 to 45 weeks GA (287–315 days), reporting an accuracy of 63% correct within 1 week and 82% correct within 2 weeks. These findings are comparable to our results, however, the maturational EEG characteristics are quite different between preterm and term neonates (Vanhatalo and Kaila, 2006; André et al., 2010).

There are two main sources which contribute to the variability seen between the automated estimate of the EMA and the known GA: biological and technical. There is biological variability between the GA and MA. The GA, used as the surrogate measure of MA, is fixed to the LMP of the mother (Engle et al., 2004). The EMA is, however, more closely related to conceptional age and conception can vary from 8 to 20 days after the LMP (Geirsson, 1991). Approximately 60% of the EMA estimates fall within ±1 week of the GA (the EMA estimator inherently caters for the 2 week bias between GA and conceptual age). There is additional biological variability that affects the developmental trajectories of each neonate. Technical variability can be introduced by the EEG recording quality, the robustness of feature extraction methods and the training process of the SVR.

In our dataset, relative beta power had the highest correlation with GA. This is not commonly investigated in studies of maturation as power in the beta band is less obvious in the preterm EEG. In fact, the absolute beta power can be considered as an estimate of the noise level of the EEG recording. There is, however, some precedence for the usefulness of relative beta power as similar features have been shown to correlate with maturation (Holthausen et al., 2000; Tolonen et al., 2007). The increase in relative beta power is most likely caused by decreasing low frequency power in the delta band (Vanhatalo et al., 2005) that is further...
accentuated by increases in absolute beta power with maturation (Tolonen et al., 2007). These findings support a flattening of the EEG frequency spectrum with early brain maturation, a trajectory that continues beyond the neonatal period (Vanhatalo et al., 2005; Fransson et al., 2013; Chu et al., 2014).

The majority of high correlating features were measures of the extremes of the distribution of EEG amplitude in the recording. There is evidence, from other quantitative analyses, that measures of amplitude have the strongest association with EEG maturation (Holthausen et al., 2000; Niemarkt et al., 2011; O’Reilly et al., 2012). This is interesting as EEG amplitude is not used in the conventional visual interpretation of the EEG which relies on estimating features of temporal EEG behaviour such as interburst interval (inter-SAT duration) and burst/SAT duration, or specific waveforms and interhemispheric synchrony (André et al., 2010; Aminoff, 2012). This may be due to physical differences in neonatal and adult EEG. In neonates, the EEG is highly focal and not, therefore, strongly affected by volume conduction as seen in adult EEG (Odabaee et al., 2013). This implies that neonatal EEG amplitude may be a more accurate measure of underlying EEG cortical activity and may reflect changes in the spatial constellation of EEG source activity, which arise from the maturation of cortex/subplate activities, cortical folding, as well as intra-cortical networking (Vanhatalo and Kaila, 2006; Kostović and Judas, 2010; Kilb et al., 2011; Iyer et al., 2015). Measures of amplitude summarised over time are also not entirely independent of temporal changes of the EEG and the presence of specific waveforms such as temporal theta and delta waves will alter relative spectral measures of amplitude with respect to frequency bands.

Our best performing EMA estimate was based on the amplitude, spatial and temporal characteristics of the EEG, although this combination was not significantly better than an estimate based on only amplitude features. This suggests that temporal and spatial characteristics provide minimal support to the EMA estimator. A point which is supported by the fact that spatial and temporal based characteristics assessed independently provided a poor estimate of EMA in the very and extremely premature neonate. The EMA estimator was improved when EEG activity state was taken into account and when multiple estimates from each neonate were averaged.

A limitation of this study is the relatively small amount of data used to optimise the EMA estimator (three 1 hour epochs from 49 neonates), even though the current database represents a significant effort of data acquisition (approximately 2 years of collection from a maternity services that averages 9000 deliveries per year). Furthermore, current research protocols in our NICU are limited to the collection of EEG from very and extremely premature neonates less than 32 weeks GA. When generating the database, we opted for more subjects rather than more epochs from each subject. In order to keep the number of epochs equal from each neonate we were limited by the neonates with the shortest, ‘predominantly artefact free’, recording duration which in this case was 3 hours. We aim to expand the number of neonates, number of epochs and the range of GA in the future. A dataset with a higher number of EEG epochs taken from a larger cohort of neonates over a greater range of EMA has potential for improving feature selection and the training of the EMA estimator. An additional relative limitation is that our benchmark for EEG normality was based on the visual interpretation by an expert. We have not included measures such as neurodevelopmental outcome or the results of imaging. Prior studies have used normal neurodevelopmental outcome or absence of neurological deficits as indirect signs of EEG normality (Holthausen et al., 2000; Niemarkt et al., 2010). While theoretically intriguing, these measures cannot exclude abnormality in the EEG, just as the presences of acute illness at the time of EEG recording does not always imply EEG abnormality. In essence, visual interpretation of the EEG is the only currently available method of determining EEG normality. Indeed, the use of automated and objective measures as developed in the present work hold promise for creating such normative, quantitative criteria when applied to large well defined EEG datasets.

Further improvements will aim to reduce both interpatient and intrapatient variability. While general improvements in estimator performance will be reflected in both forms of variability, there may be features that improve one aspect of the overall variability over the other. Improvements in inter-patient variability require the development of EEG features that are better correlated with the MA of the brain. Features such as complexity and inter-connectivity are potentially applicable (Stevenson et al., 2007; Janjarasjitt et al., 2008; Tokariev et al., 2012; Koolen et al., 2014; Meijer et al., 2014). Improvements in intra-patient variability require features that are robust to changes in time of recording after birth and EEG recording environment. Features such as signal to noise ratio, impedance and others that respond to the presence of several artefacts are potentially applicable (Stevenson et al., 2014b). Once the development of the EMA is finalised it must be validated on a cohort of preterm neonates with normal and abnormal EEG for age across a range of aetiologies. This will determine if the EMA estimator can generate a measure of dysmaturity that correlates with the presence of abnormality in the EEG; a correlation that will be clinically useful.

5. Conclusion

We developed an automated method of estimating functional brain maturation in very and extremely premature neonates based on analysis of the EEG. The estimator generated an EMA that was within 2 weeks of GA in 87% of all EEG epochs in our development dataset. Features of EEG amplitude had the highest correlation between EMA and GA. The EMA estimator was improved by targeting feature extraction to periods of low EEG activity (quiet sleep) and averaging estimates across multiple epochs per neonate. The proposed EMA estimator is an important first stage in the development of a novel automated EEG maturity index for use as a neurological monitoring tool in the NICU. It is also provides a simple summary measure of a complex interpretation of the EEG.

Acknowledgements

The authors acknowledge the support of Science Foundation Ireland (12/IP/1369). Data collection was supported by the Wellcome Trust UK (085249). JMOT was supported by the Irish Research Council (GOIPD/2014/396). GBB was supported by Science Foundation Ireland (12/RC/2272) and the Wellcome Trust UK (098983). SV was supported by the Academy of Finland (276523 and 288220), and the Sigrid Juselius Foundation. NJS was supported by Science Foundation Ireland (12/IP/1369) and the EU Framework Programme for Research and Innovation (H2020-MCSA-IF-656131).

Conflict of interest: None of the authors have potential conflicts of interest to be disclosed.

Appendix A

The EMA estimate equation for the ith hour long epoch of EEG which is summarised by a feature vector \( x_i \) containing \( m \) features, is defined as,

\[
y_i = \left( \sum_{j=1}^{m} (x_{pi} - x_{pi})K(x_p, x_i) \right) + b
\]  

(A1)
where $\hat{y}_i$ is the EMA estimate, $(x_1, x_2)$ are weights derived from Lagrangian multipliers, $x_p$ is a $m$ by $p$ matrix containing $p$ support vectors, which are a subset of the training data, $K(x, x)$ is the kernel function, and $b$ is a bias term (Smola and Schölkopf, 2004). In the case of linear SVR, the kernel function is the inner product between the feature and support vector. The EMA estimation equation can, therefore, be written as a standard linear predictor,

$$\hat{y}_i = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m + b$$

where $a = \sum_{p=1}^{p} (x_p - \mathbf{x})^2$ is a vector of length $m$.

References


