DEPRECATED: pre-2009 .ERR format

Note: As by build 20090729 of the Flux Simulator, the .ERR format has changed to allow for both, quality-based models and such without. The article here describes the .ERR format before build 20090729.
Error model (ERR) files are proprietary to the Flux Simulator and optionally used during the sequencing process. Data is organized in blocks and presented in tokens separated by white-spaces. There are 4 different block types:

- Model Pool Summary (one per file)
- Crosstalk Table (one per file)
- Position-Error Models
- Sequence-Error Models

Probability distributions over a discrete value space (e.g., quality values, substitution symbols, etc.) are coherently described by their cumulative distribution functions (CDFs). As by their nature, the number series in a CDF have to be monotonously increasing with (at least) the last value of a series being 1.

Model Pool Summary

```
#MODEL minQual maxQual tholdQual nrInstances
[p(minQual) p(minQual+1) ... p(maxQual-1) p(maxQual)]
```

Expression (Example)	Explanation
#MODEL	tag introducing the model description block
minQual (-40)	minimum quality: the minimum value for qualities in the described error models. Currently exclusively integer quality models (as Illumina and phred qualities) are addressed. Therefore, subsequent CDFs over quality spectra have all the length (maxQual - minQual + 1). **Only** for error files that have been built with quality values.
maxQual (40)	maximum quality: highest value of the quality spectrum, an integer - see above. **Only** for error files that have been built with quality values.
tholdQual (.)	the threshold quality: level below which below which all base-calls have been considered "problematic" or "accident", regardless whether the corresponding base had been called correctly or not. If none such threshold has been applied, tholdQual should be set to "." **Only** for error files that have been built with quality values.
nrInstances (916311)	number of instances: on how many observations (i.e., reads) the error model has been estimated on
p(minQual), … , p(maxQual)	CDF over qualities of "unproblematic" base calls. A base call is considered as unproblematic iff it is (i) correct and (ii) equal or above the level specified by tholdQual. **Only** for error files that have been built with quality values.

Crosstalk Table

```
#CROSSTALK letter
[minQual] p(A) p(C) p(G) p(N) p(T)
[minQual+1] p(A) p(C) p(G) p(N) p(T)
...
[maxQual-1] p(A) p(C) p(G) p(N) p(T)
[maxQual] p(A) p(C) p(G) p(N) p(T)
```

Expression (Example)	Explanation
#CROSSTALK	tag that introduces a crosstalk description block
letter (A)	Symbol, for which the crosstalk is specified as the observed substitution rates broken down by quality levels.
minQual … maxQual (-40,…, 40)	quality level for the following observed substitution rates p(X) apply. **Only** for error files that have been built with quality values.
p(A),p(C),p(G),p(N),p(T)	probabilities (or CDF) for the symbol specified by letter to be substituted by A, C, G, N, or T.

Position-Based Error Models

```
# PositionErrorProfile start length baseProb
[start p(minQual) p(minQual+1) ... p(maxQual-1) p(maxQual)
(start+1) p(minQual) p(minQual+1) ... p(maxQual-1) p(maxQual)
...
(start+length-1) p(minQual) p(minQual+1) ... p(maxQual-1) p(maxQual)]
```

Expression (Example)	**Explanation**
#PositionErrorProfile	tag that introduces position error profile block
start (26)	first position in the read affected by this error model ~~(1-based)~~ **(0-based)**
length (11)	extension of the "problem" captured in this error profile. Consequently, the 0-based index of the last position affected is (start+length-1).
baseProb (6.875394925958544E-5)	probability as fraction of reads that shared this problem in the observed dataset. Multiplying this probability with the value nrInstances in the #MODEL block recasts the number of instances in which this error has been observed.

Sequence-Based Error Models

(forthcoming)