

QUALIDADE DOS DADOS DIÁRIOS DE CHUVA NO ESTADO DE SÃO PAULO

Airton F.S. Xavier

Depto. Estat. Matem. Aplic., Univ. Fed. Ceará

Teresinha de Ma. Bezerra S. Xavier

Depto. de Geografia /Mestrado em Des. e Meio Ambiente, Univ. Fed. Ceará

Maria Assunção F. da Silva Dias

Depto. Ciências Atmosféricas, IAG-USP

José Roberto C. de Toledo

DAEE-CTH e E.E.U.M. S. Paulo

E-Mail: alxavier@ia.ufc.br

RESUMO

Neste artigo analisamos o problema da qualidade de dados diários de chuva no Estado de São Paulo do banco de dados do DAEE-CTH. Para as séries analisadas, sérios problemas são detectados até 1970; já de 1972 a 86, em geral, os dados podem ser considerados como tendo boa qualidade; porém, a partir de 1988, detecta-se neles certa deterioração, embora não sejam tão seriamente afetados quanto ocorria no primeiro período. Em consequência disso, recomenda-se uma análise preliminar e revisão de toda e qualquer série de dados diários de chuva que deva ser utilizada numa pesquisa, especialmente quando se referir ao contexto de estudos sobre mudanças climáticas.

INTRODUÇÃO

Nas últimas décadas tem-se dado ênfase no estudo de mudanças do clima, no âmbito de possíveis influências de caráter local ou global. Aí, as variáveis envolvidas referem-se à pluviometria, à temperatura do ar, a vazões de rios ou níveis de lagos, etc. Nesse tipo de estudo devem ser impostas, necessariamente, exigências estritas quanto à fidedignidade dos dados, sob pena de sérias dúvidas quanto aos resultados e conclusões.

Aqui, nos restringimos ao caso da chuva e à qualidade dos seus dados. Lembremos que a maioria dos trabalhos sobre alterações na pluviometria executados em várias regiões do globo tem sido centrada no estudo de séries anuais ou mensais. Porém, é muito difícil que se consiga detectar mudanças na pluviometria decorrentes, por exemplo, dos processos de urbanização ou de industrialização, às custas exclusivamente de uma análise da variação de alturas da chuva a nível anual ou mesmo mensal; com efeito, dessa maneira, não ficam separadas ou discriminadas as ocorrências de mínimas ou de máximas intensidades da precipitação que, justamente, se espera possam ser os eventos que constituem as respostas específicas aos processos mencionados.

Uma análise mais fina, calcada em dados diários, horários ou mesmo referentes a intervalos de tempo ainda mais curtos, sem dúvida poderia oferecer condições para que se obtenham conclusões e interpretações menos controvertidas. Porém, é rara a disponibilidade de séries diárias, não apenas suficientemente longas, bem como, confiáveis e também apresentando poucas falhas; para não falar de dados horários ou em intervalos de minutos.

Uma dificuldade suplementar decorre da circunstância de que na pesquisa climática há em geral a necessidade de comparar como se comportam séries pluviométricas de

localidades vizinhas. Por exemplo, as que se situam numa área urbana e outras em áreas rurais próximas; ou então, no caso de localidades a jusante e a montante de uma fonte poluidora, com respeito ao fluxo do vento. Ora, como se percebe, o uso de técnicas para "completamento de falhas" torna-se em tais casos um procedimento certamente inadequado, principalmente que isso implicaria na contaminação dos dados pluviométricos de uma localidade pelos dados das localidades vizinhas, o que viria a falsear a análise. Tal advertência se aplica, inclusive, no caso do emprego de métodos de análise objetiva destinados à estimativa e síntese de dados diários sobre uma grade, que são válidos apenas quando destinados às finalidades muito específicas para as quais foram desenvolvidos (para a obtenção de normais climáticas diárias ou como mera etapa operacional no contexto da previsão meteorológica).

Em trabalho anterior, no qual foram utilizados os dados diários de chuva da *estação meteorológica do IAG/USP* (posto E3-035) situada no Parque Estadual das Nascentes do Ipiranga (Parque do Estado), na cidade de São Paulo (Água Funda), para o período 1933-86, havíamos documentado certas características marcantes na evolução da chuva diária nesse local, como a atenuação progressiva das precipitações diárias mais fracas (< 5 mm) e o reforço daquelas mais intensas (a partir de 30 mm); ver Xavier et al (1994a). Os mesmos resultados foram comprovados ao se estender a análise para o período 1933-93, conforme Xavier et al (1994b).

Foi ainda tentada a comparação dos resultados obtidos com a série do IAG/USP, com respeito a outros postos situados numa área bastante ampla, distribuídos aproximadamente segundo um transepto na direção SE-NW, indo desde o litoral até o município de Campinas e passando pela cidade de São Paulo; além de terem sido incluídos outros postos à direita e à esquerda desse transepto. Resultados preliminares,

nesse sentido, foram também apresentados em Xavier et al (1994b).

Em particular, apareceram indícios de anomalias com respeito a alguns postos situados numa área sob a possível influência da poluição industrial oriunda do *município de Cubatão*, que se traduziam como "rupturas" ocorridas nos dados, relativamente às chuvas diárias menos intensas (< 5 mm). Como, por exemplo, foi verificado para o posto E3-037 (Paranapiacaba/PRP) situado no alto da Serra do Mar, a cavaleiro do vale onde se situa o parque industrial de Cubatão. Contudo, as "rupturas" eram muito abruptas, conforme se vê na Figura 1, assim, dificilmente explicáveis somente em termos do efeito da poluição atmosférica; de fato, não há notícias de aumentos também subitâneos dos níveis de poluição que pudessem explicar tais "rupturas". Portanto, resolveu-se examinar a possibilidade de que decorressem de erros ou de inconsistências nos dados, o que veio a ser comprovado, conforme veremos. Resultados preliminares e resumidos desta pesquisa foram apresentados em Xavier et al (1995) (1996).

DADOS E METODOLOGIA

Consideramos a região entre 46,6 graus ($46^{\circ} 36'$) W e 46,0 graus W; e entre 23,7 graus ($23^{\circ} 42'$) S e 24,0 graus S. No interior desta área são contados cerca de 40 (quarenta) postos cujos arquivos da chuva diária foram obtidos do DAEE (Depto. de Águas e Energia Elétrica) - CTH (Centro Tecnológico de Hidráulica e Recursos Hídricos). Levando em conta um critério de localização com respeito à área problema de Cubatão e também eliminando aqueles para os quais existiam muitas falhas, foram selecionados 10 (dez) postos, cujas séries da chuva diária passaram a ser examinadas, acuradamente.

Tem-se, na Figura 2, um esboço em 3D da topografia da região, com uma localização desses postos; aí, o relevo (em termos da altitude) está representado

exageradamente com respeito às outras duas dimensões geográficas, as isolinhas consecutivas correspondendo a diferenças de 25 metros. Pela inspeção dessa figura, podem ser claramente percebidos os dois postos E3-101/CUB (Cubatão, 6 m) e E3-038/PIA (Piaçaguera-Cubatão, 5 m), situados no vale da cidade industrial de Cubatão, separados dos postos E3-040/UIT (Usina Itatinga-Bertioga, 10 m) e E3-106/BER (Bertioga, 3 m), a nordeste, por um contraforte da serra que atinge altitudes até cerca de 450 m. Sobre este contraforte encontra-se o posto E3-041/CAE (Caeté-Santos, 200 m). Quanto aos quatro postos E3-037/PRP (Paranapiacaba-S. André, 800 m), E3-042/RIT (Represa Itatinga-Bertioga, 720 m), E3-104/TPL (Terceiro Plano da S. Nova-Cubatão, 670 m) e E3-109/ALS (Alto da Serra-S. Bernardo do Campo, 760 m), estão localizados no bordo superior do paredão da Serra do Mar; finalmente, tem-se o posto E3-143/VIA (Via Anchieta-Cubatão, 400 m), a meia encosta e a cavaleiro do vale de Cubatão.

Por motivo da dúvida sobre a qualidade desses dados, escolheu-se para fins comparativos uma série temporal de observações diárias da chuva, suficientemente extensa, sem falhas e, sobretudo, confiável. Tal escolha recai sobre a estação meteorológica do IAG/USP/E3-035, em razão dos argumentos expostos em Xavier et al (1994a); ou seja, são dados diários ininterruptos para todo o período 1933-93, submetidos a rigoroso controle de qualidade e, assim, considerados fidedignos. Para a estação meteorológica do IAG (Instituto Astronômico e Geofísico), foram utilizados os dados fornecidos por este órgão da Universidade de São Paulo. De fato, tais dados diários originam-se de observações horárias, acumuladas no período de 0 horas às 24 horas de cada dia. A mesma série, com a sigla E3-035, está também disponível no banco de dados do DAEE-CTH, porém com as observações horárias acumuladas das 7 sete horas de cada dia às 7 horas da manhã do dia seguinte. Evidentemente, os valores das duas séries não coincidem; contudo, conforme

testes por nós desenvolvidos no contexto da pesquisa anterior [Xavier et al (1994a) (1994b)], ambas se comportam de forma similar, estatisticamente, pois conduzem a resultados entre si comparáveis, como aliás seria de esperar.

No que se refere à análise da qualidade dos dados, tomamos como critério a inspeção dos histogramas de frequências para a chuva diária, nos períodos: (i) até 1970; (ii) de 1972 a 86; (iii) a partir de 1988. A escolha desses períodos foi ditada pelas "rupturas" observadas com respeito às chuvas menos intensas (< 5 mm); "rupturas" que ocorreram entre 1970 e 1972, e também entre 1986 e 88, conforme a Figura 1. Para confeccionar os histogramas foram considerados apenas os valores não nulos da altura da chuva e, ademais, com respeito à maioria dos postos, com um truncamento para os valores acima de 50 mm.

Desprezar os valores nulos, de fato, corresponde a descartar a parte da distribuição de probabilidades concentrada na origem, restando apenas a sua parte "contínua" (ou seja, como será visto adiante, consideramos a chuva diária X sob a condição $X > 0$).

Note-se que para os postos pluviométricos na área de influência de Cubatão, utilizados neste trabalho, o valor $x=50$ mm corresponde, aproximadamente, ao quantil $q(0,90)$; donde se conclui que, nestes casos, o truncamento acima de tal valor corresponde a descartar valores extremos da chuva diária numa proporção cerca de 10% de todas as observações não nulas. A rigor, não há uma razão de caráter teórico para justificar a escolha do quantil $q(0,90)$. Em princípio, o truncamento possui eminentemente uma finalidade prática, que consiste em não deixar uma cauda muito longa no gráfico, que dificultaria a visualização de possíveis irregularidades no histograma. No caso dos dados de chuva da estação pluviométrica do IAG/USP, sem dúvida, a escolha de um truncamento em $x = 20$ mm seria o mais apropriado, no mesmo contexto, pois este é o valor aproximado do quantil $q(0,90)$ para este

posto (na Baixada Santista e faldas da Serra do Mar, os índices pluviométricos são bem mais elevados do que no planalto).

Lembremos que histogramas para a chuva diária costumam apresentar forte assimetria (assumindo a forma clássica de *J-invertido*). Assim, para a modelagem da variável aleatória $X = \text{"chuva diária"}$, sob a condição $X > 0$, preconiza-se o emprego das leis gama ou exponencial modificada; como em Woolhiser & Roldan (1982). Na falha de tais modelos, outros têm sido propostos, como é o caso do modelo hiperbólico referido por Subbaramayya & Rao (1964); ou um modelo especial que permite relacionar, entre si, as porcentagens acumuladas da chuva diária e as dos números de dias de chuva, como em Ananthakrishnan & Soman (1989), Soman & Kumar (1990) e Kutiel (1990). Lembramos, ainda, a possibilidade do emprego do modelo "kappa" proposto por Mielke (1973); ou o modelo "hiper-gama" de Suzuki (1964-65). De fato, considerar o valor X da chuva diária condicionalmente a $X > 0$, significa que são deixados à parte os valores nulos da chuva diária (dias sem chuva). A introdução dos valores iguais a zero obrigaria a considerar um modelo misto, caracterizado por uma probabilidade $p > 0$ centrada no ponto $X = 0$, conforme refere Juras (1994); aliás, um tal modelo misto, envolvendo a lei gama, havia sido tratado por Wilks (1990).

Sem dúvida, o ajuste de um desses modelos para as chuvas diárias e, em cada caso, a subsequente mensuração de como a densidade empírica se ajusta à densidade teórica pode, decerto, oferecer algum interesse. Mas adiantamos que seria, apenas, um procedimento complementar no contexto do nosso problema, porém incapaz de nos propiciar, exatamente, os mesmos informes que são extraídos a partir da inspeção dos histogramas.

Assim, tivemos a idéia de insistir na inspeção dos histogramas (densidades empíricas) como o procedimento de eleição para avaliar a qualidade dos dados diários de chuva. Como veremos, é um método muito simples, mas que conduz a resultados válidos,

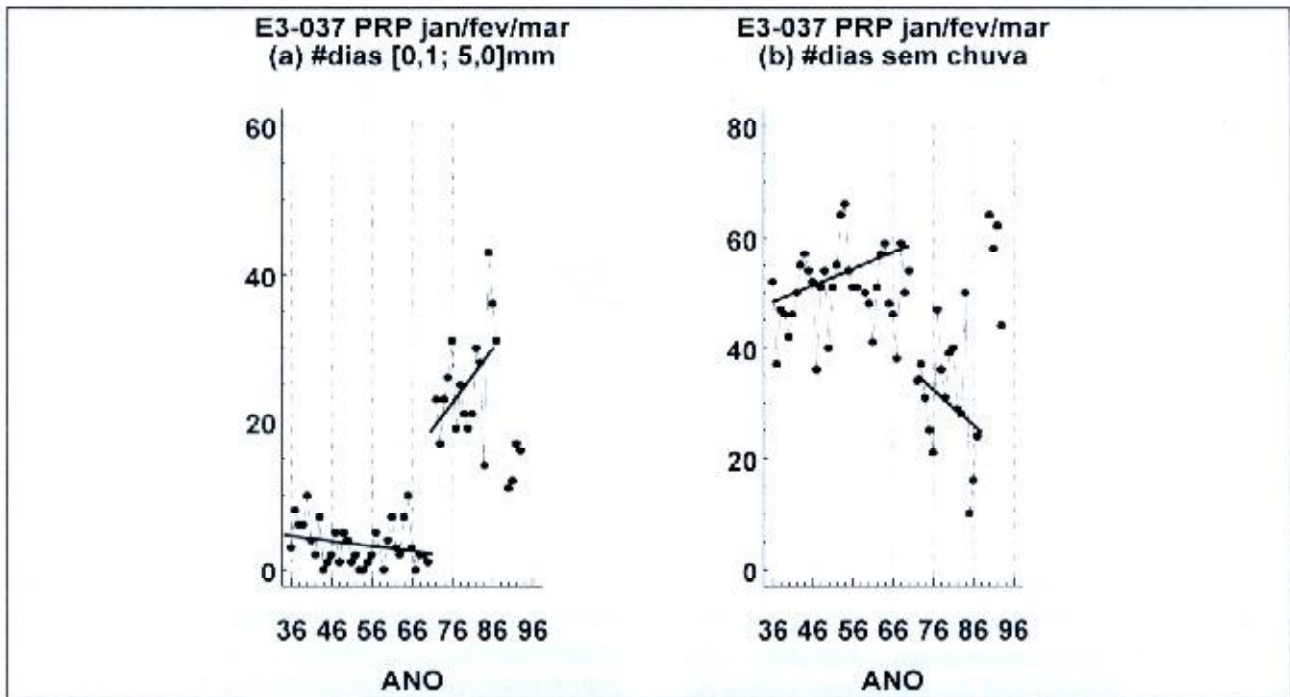
em particular no que se refere à identificação dos vícios nos dados. De fato, posteriormente, encontramos a mesma idéia brevemente exposta num artigo de Lavery et al (1992), com respeito à análise de um conjunto histórico de dados de chuva, para a Austrália.

Esses pesquisadores, igualmente, sugeriram o emprego da técnica que consiste em analisar os desvios (principalmente os que ocorrem nas caudas) no gráfico em que a "curva acumulativa" é linearizada através de uma transformação adequada, como é o caso do tradicional emprego do "papel probabilístico" que transforma em uma reta a curva acumulativa da lei normal. Mas se trata de um procedimento, complementar no nosso contexto, que exige seja conhecido o modelo probabilístico apropriado, em cada caso. Outra técnica análoga que provavelmente poderá ser explorada, nessa mesma direção, é o uso de "Q-Q plots" (ou "quantil-quantil plots"), a cujo respeito o leitor poderá consultar Easton & McCulloch (1990) e a bibliografia por eles coletada.

Omitimos os detalhes a respeito da metodologia para a regressão multifásica, utilizada "en passant" no caso da Figura 1. Para a regressão bifásica (ou em dois pedaços), o leitor interessado é remetido a Hinkley (1969, 1971), Freeman (1984) e Solow (1987). Contudo, a contrário desses autores, não impomos uma condição de "continuidade"; ou seja, as retas de regressão consecutivas podem apresentar "rupturas" propriamente ditas e, não apenas, "dobraduras" ou "inflexões".

RESULTADOS

Nesta seção apresentamos os resultados concernentes à análise dos histogramas, com respeito aos vários postos pluviométricos (ver a Figura 2) e aos diversos períodos de tempo. A regularidade ou irregularidade dos histogramas serve de indicio para a boa qualidade dos dados ou



(a) gráfico da variável p05 = # dias com chuva < 5 mm, no trimestre janeiro/fevereiro/março, de 1936 a 1993

(b) gráfico da variável n0 = # dias sem chuva, no mesmo trimestre e ao longo do mesmo período de anos (observam-se "rupturas", especialmente nítidas entre os anos 1970 e 1972, mas que são artefatos resultantes de erros ou inconsistências nos dados)

Figura 1. Posto E3-037/Paranapiacaba.

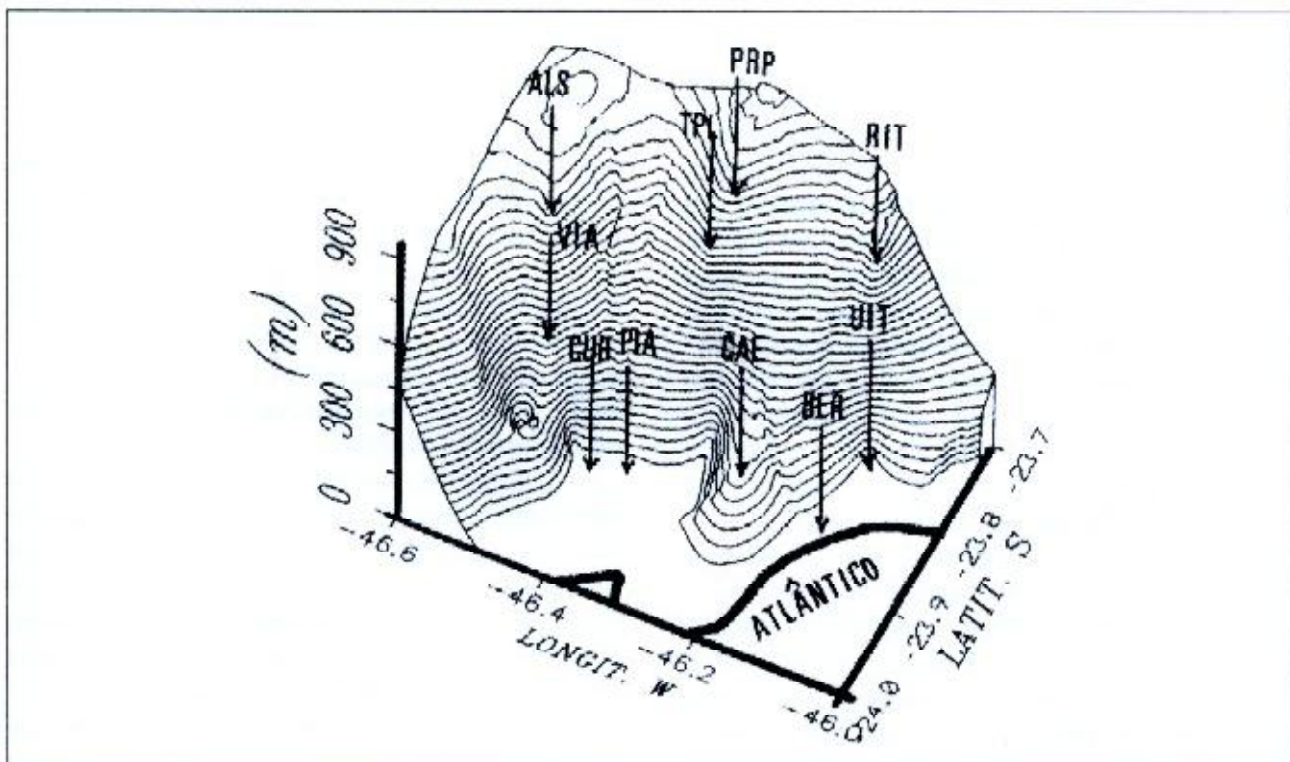


Figura 2. Esboço tridimensional da área estudada, com a situação dos 10 (dez) postos pluviométricos.

para a existência de erros ou de inconsistências, nos mesmos. Por outro lado, quando ocorreram irregularidades flagrantes, foram inspecionadas as planilhas contendo as anotações originais feitas pelos observadores, no sentido de tentar complementar a identificação dos vícios existentes nos dados ou de confirmar sua natureza.

Dados diários dos postos E3-035/IAG/USP e E3-104/ALS

Na Figura 3 temos os histogramas para os dados diários desses dois postos, nos três períodos: (i) até 1970 ; (ii) 1972-86 ; (iii) 1988-92 ou 93. No caso do IAG/USP, os histogramas sofrem um truncamento acima de $x=20$ mm, pelo motivo já explicitado anteriormente. Porém, são também apresentados os histogramas para seus dados aproximados para décimos de polegadas (tem-se $0,1 \text{ pol} \approx 2,54 \text{ mm}$) e em seguida retransformados para milímetros. A razão para isso será melhor compreendido mais adiante; neste último caso, o truncamento foi mantido em $x=50$ mm, por motivo óbvio que ressalta dos histogramas respectivos. Note-se, para os dados do IAG/USP nos três períodos, que os histogramas dos dados diários apresentam a forma clássica de *J-invertido*; portanto, tendo em vista sua regularidade, trata-se de uma evidência da boa qualidade dos dados. Contudo, mínimas irregularidades observadas em alguns desses histogramas podem ser atribuídas a flutuações de amostragem; especialmente, com respeito ao último período (desde 1988), cuja amostra envolve 888 observações contra 8110 no primeiro período. Quanto ao posto E3-104/ALS, também não há indícios de problemas sérios nos dados, a partir da inspeção dos histogramas. Assim, para os demais postos a serem analisados, a seguir, um indicio ou evidência da boa qualidade dos dados seria o comportamento, regular, dos histogramas respectivos ou das curvas de frequências.

Dados diários dos postos E3-037/PRP, E3-038/PIA, E3-041/CAE e E3-104/TPL

Na Figura 4 temos os histogramas para os postos E3-037/PRP, E3-038/PIA e E3-041/CAE, com respeito aos mesmos três períodos, quando observamos a existência de irregularidades para cuja explicação exigir-se-á uma análise criteriosa.

No caso dos postos E3-037/PRP (Paranapiacaba) e E3-038/PIA (Piaçaguera), deve-se notar que ambos foram operados pela empresa ferroviária inglesa "São Paulo Railway", até 1946, data de sua encampação pelo governo federal. Contudo, há ainda de se ressaltar que continuaram em operação nestes dois postos, até dezembro de 1970, os pluviômetros ingleses graduados em polegadas (*pol*), que somente a partir de 1971 foram substituídos pelos pluviômetros padronizados em milímetros (*mm*), da rede paulista. A transformação para milímetros dos dados originais em polegadas (certamente medidos com aproximação até $0,1 \text{ pol}$) explica a ocorrência de valores múltiplos de $\approx 2,54$ mm (valor de um décimo de polegada inglesa); o que vai implicar na aparente "discretização" dos histogramas para esses dois postos, no período 1936-70. Por outro lado, daí decorre uma perda de informação com respeito às chuvas de intensidades muito pequenas, que acabam computadas como correspondentes a alturas nulas. Nessa perda de informação reside o mecanismo responsável pelas aparentes "rupturas" verificadas na Figura 1, em torno de 1970-72, que assim se revelam de forma inequívoca serem meros artefatos.

Contudo, esse não é o único problema com esses dados, no período 1936-70, diante das grandes irregularidades detectadas nos histogramas respectivos. Note-se que, na Figura 3, havíamos introduzido de maneira artificial uma "discretização" nos dados diários de chuva do IAG/USP ; mas, não obstante, os histogramas mantiveram uma regularidade notável em termos do seu aspecto geométrico.

Na verdade, relativamente aos postos E3-037/PRP e E3-038/PIA, além da ocorrência (até 1970) de muitos valores múltiplos e sub-múltiplos de 25,4 mm (valor de uma polegada inglesa), a análise minuciosa das folhas originais de observação (anexadas às folhas com os dados transcritos em milímetros), revela também a existência de muitos valores iguais a 0,5 pol (e sub-múltiplos tais como 0,1 e 0,2 pol). Deve-se observar que a tais constatações já haviam chegado os engenheiros do DAEE-CTH que se incumbiram, durante os anos de 1970-1971, da análise e publicação dos dados pluviométricos do Estado de São Paulo medidos até 1970. Nesta época, embora na ausência de provas irrefutáveis de natureza estatística, já haviam decidido pela não publicação dos dados em questão, por terem sido julgados e classificados como não fidedignos. Assim, a análise estatística agora empreendida vem corroborar a decisão tomada naquela ocasião. Ou seja, de que os dados diários dos postos E3-037/PRP e E3-038/PIA (como também os do posto E3-104/TPL) no período até 1970, são inteiramente suspeitos e precários.

Como não existem informações sobre a natureza dos pluviômetros usados (isto é, pluviômetros ingleses), não se sabe se sua precisão seria de fato compatível com medidas fracionárias de polegada, ou se os vícios nas medidas anotadas decorreriam simplesmente da negligência ou incompetência dos observadores, que eram funcionários da "São Paulo Railway". Pode-se, também, aventar a hipótese de que houvesse instruções aos observadores, por parte dos engenheiros ferroviários, no sentido de não de preocuparem com chuvas pequenas, que não causam danos às obras de arte da ferrovia (tais como aterros, pontes, viadutos e bueiros). Mas é indubitável que esses dados acham-se eivados de erros de medição, que os prejudicam de maneira irremediável.

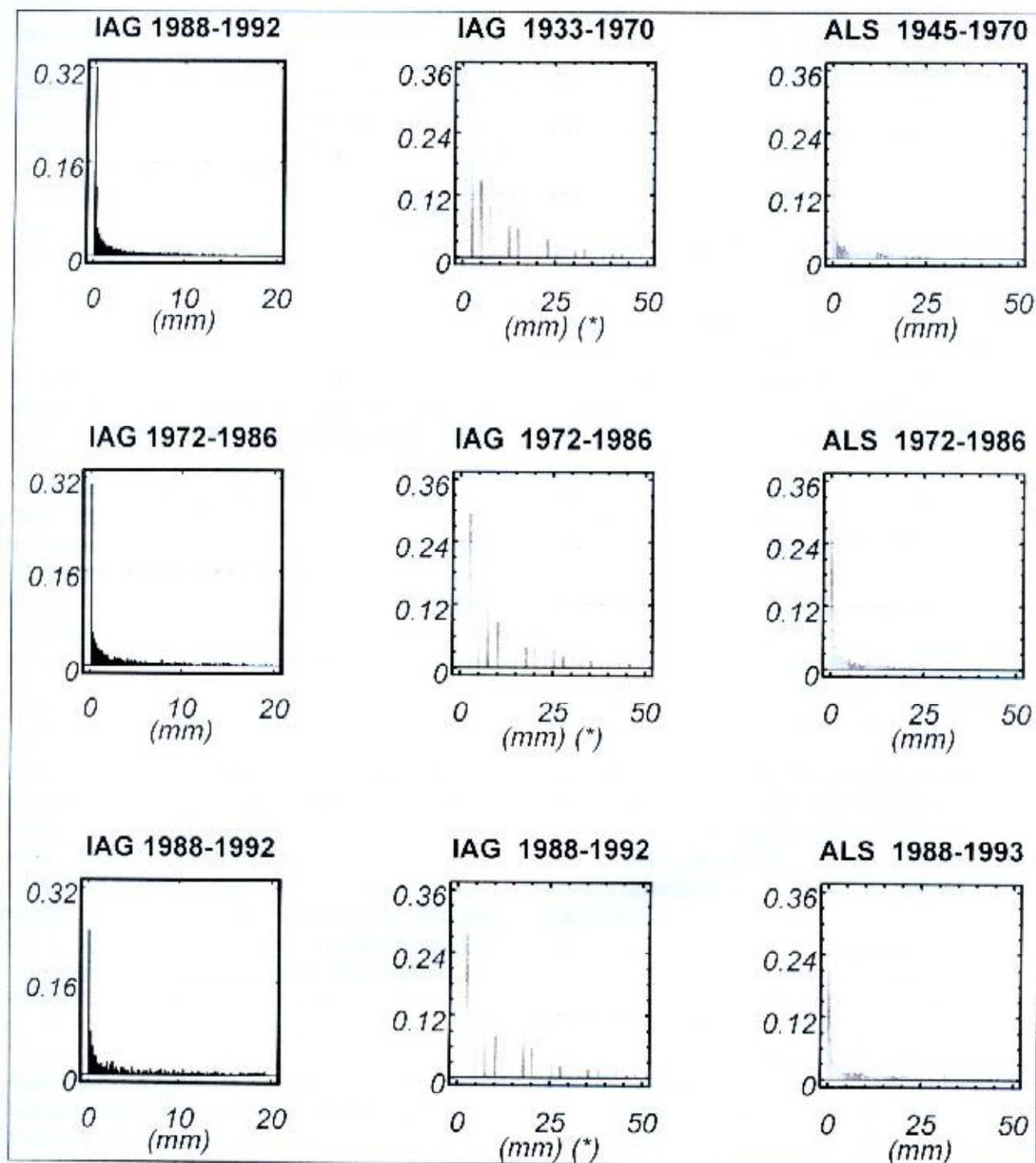
Com respeito ao posto E3-041/CAE (Caeté-Santos), no período 1938-70, não ocorre evidentemente o fenômeno de "discretização" (originado apenas quando do

uso de pluviômetros calibrados em polegadas), porém ainda são denotadas grandes irregularidades; em particular, ficando evidenciado o aparecimento de valores anômalos correspondentes a múltiplos de 5 mm ou de 10 mm. Um exame mais detalhado dos dados, a partir da inspeção das planilhas originais, vem, neste caso, revelar que tais valores anômalos decorrem de erros de "arredondamento" das leituras (seja pelo abandono da parte decimal, seja pela consideração apenas de "números de provetas cheias", o que corresponde a anotações de múltiplos inteiros de 10 mm).

No período 1972-86, pelo contrário, a regularidade dos histogramas para esses vários postos pluviométricos revela a boa qualidade dos dados anotados, principalmente no caso do posto E3-037/PRP. A bem da verdade, para E3-038/PIA, ainda são denotadas algumas irregularidades, mas que seriam provavelmente melhor percebidas e avaliadas usando um truncamento acima de outro valor (em vez de 50 mm) para a confecção do histograma.

Finalmente, a partir de 1988, há indícios de reaparecerem outras irregularidades, embora não sejam tão graves quanto as observadas no primeiro período (i.e., até 1970). Na discussão, consideraremos razões de ordem operacional que explicariam seja o reaparecimento, seja o recrudescimento das irregularidades; mas sem que se possa afastar, mais uma vez, o possível efeito das flutuações de amostragem, especialmente quando se lida com amostras de menor tamanho.

São omitidos os gráficos para E3-104/TPL (Terceiro Plano da Serra Nova-Cubatão) que, para os períodos 1961-70 e 1972-81, se comportaram como os de E3-037/PIA, em análogos períodos. Para este posto não existe um gráfico referente ao último período, uma vez que suas observações são disponíveis apenas de 1961 até 1980.



(*) dados aproximados para 0,1 pol e retransformados para milímetros.

Figura 3. Histogramas da altura da chuva diária, para os postos E3-035 (IAG/USP) e E3-109 (ALS), nos períodos: (i) até 1970; (ii) 1972-1986; (iii) a partir de 1988.

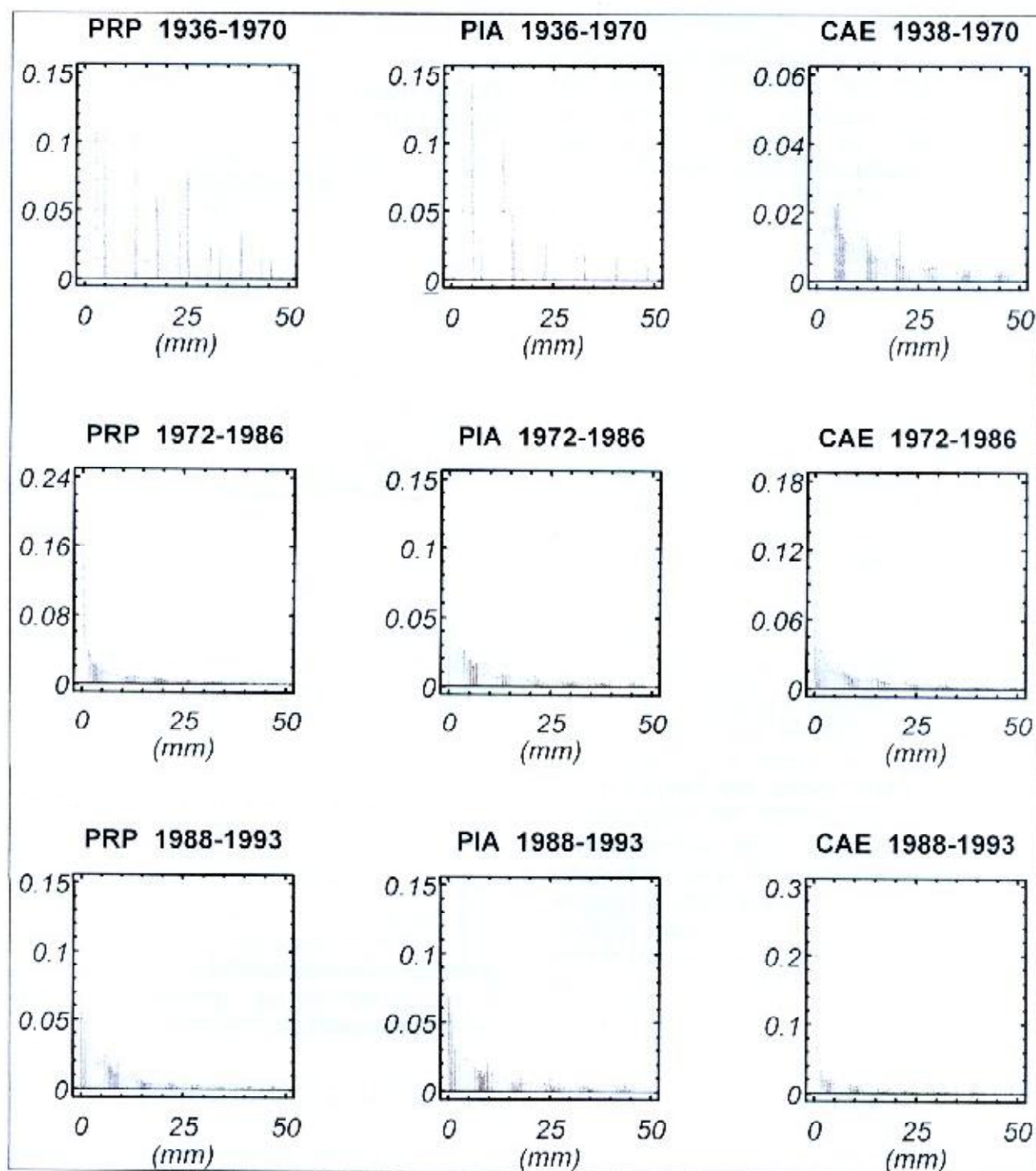


Figura 4. Histogramas da altura da chuva diária, para os postos E3-037/PRP, E3-038/PIA e E3-041/CAE, durante os períodos: (i) até 1970; (ii) 1972-1986; (iii) a partir de 1988.

Dados diários dos postos E3-101/CUB e E3-143/VIA

No caso dos postos E3-101/CUB (Cubatão) e E3-143/VIA (Via Anchieta-Cota 400-Cubatão) (ver Figura 5), embora não ocorra o problema decorrente da transformação da unidade inglesa, são contudo detectadas graves distorções dos dados diários, no período até 1970.

Como ocorria com os dados do posto E3-041/CAE, são novamente nítidos os acúmulos de valores correspondentes a múltiplos de 5 mm ou 10 mm, especialmente no caso do posto E3-143/VIA, até o ano de 1970. Assim, a inspeção e análise das anotações originais revela que, a partir de agosto de 1965, ocorreram muitos arredondamentos nos dados, mais frequentes em certos sub-períodos do que noutros. Isso se explica pela mudança de observador, que substituiu o que operava o posto desde dezembro de 1962. Tal operação defeituosa persistiu até dezembro de 1969.

Quanto ao posto E3-101/CUB, a desorganização ou irregularidade do histograma é ainda mais drástica, no mencionado período, isto é, até 1970. Curiosamente, o exame dos dados originais deste posto não revela a ocorrência de arredondamentos, apesar das irregularidades muito evidentes quando se inspeciona o histograma respectivo. Uma hipótese, porém de comprovação difícil, é que poderia ter havido (de 1956 a 1970) grande número de alturas de chuva acumuladas para dois ou mais dias e registradas como sendo de um único dia, devido à negligência dos observadores.

No período 1972-1986, passamos a denotar evidência de certa fidedignidade dos dados, especialmente para o posto E3-143/VIA. Porém, no caso de E3-101/CUB, no mesmo período, embora não mais sejam verificadas as flagrantes irregularidades do período precedente, não obstante ainda parecem existir mínimas irregularidades.

Finalmente, de 1988 a 1993, para os dois postos E3-101/CUB e E3-143/VIA, mais

uma vez denota-se certa deterioração na qualidade dos dados, diante do reaparecimento de irregularidades nos histogramas.

Dados diários dos postos E3-042/RIT, E3-040/UIT e E3-106/BER

Na Figura 6, temos os histogramas nos três períodos consecutivos, para os postos E3-042/RIT (Represa Itatinga-Bertioga) e E3-040/UIT (Usina Itatinga-Bertioga).

No caso do posto E3-042/RIT são ainda denotadas grandes irregularidades até 1970 (primeiro período). No período seguinte (1972-86), pelo contrário, o histograma é muito regular em termos de apresentar a forma clássica em J-invertido, o que denota a boa qualidade dos dados. Finalmente, a partir de 1988, também mediante a inspeção do histograma respectivo, não se denota perda de qualidade nos dados.

Quanto ao posto E3-040/UIT, até 1970, as irregularidades não parecem ser tão flagrantes; contudo, podem ser observados pequeninos "picos" correspondentes a valores como 10 mm, 20 mm, 30 mm, etc., denotando a ocorrência de aproximações espúrias aplicadas aos dados (mais uma vez, correspondendo a números inteiros de provetas cheias). Por outro lado, no período 1972-86, há indício de boa qualidade para os dados deste posto pluviométrico; bem como, no último período (a partir de 1988) não ocorre uma degradação evidente nos seus dados.

Com respeito ao posto E3-106/BER (Bertioga), são omitidos os histogramas; mas o comportamento é análogo, nos três períodos, aos dois postos anteriores.

DISCUSSÃO E CONCLUSÕES

Na secção precedente já se discutiu, com pormenores, a origem da maioria dos vícios existentes, revelados através da análise dos histogramas dos dados diários para os dez postos pluviométricos aqui considerados. Além de erros nas medidas, explicáveis por

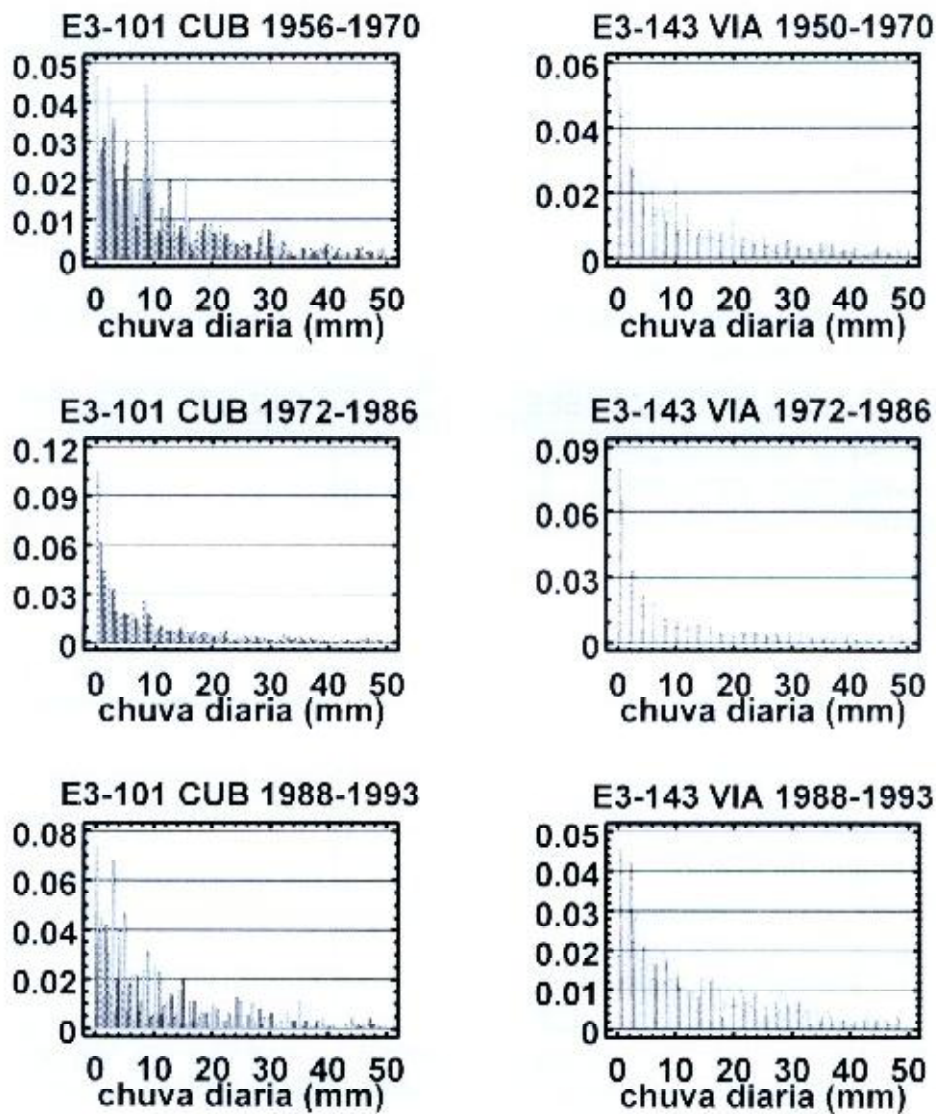


Figura 5. Histogramas da altura da chuva diária, para os postos E3-101/Cubatão e E3-143/Via Anchieta-Cota 400//Cubatão (análogos períodos das figuras precedentes).

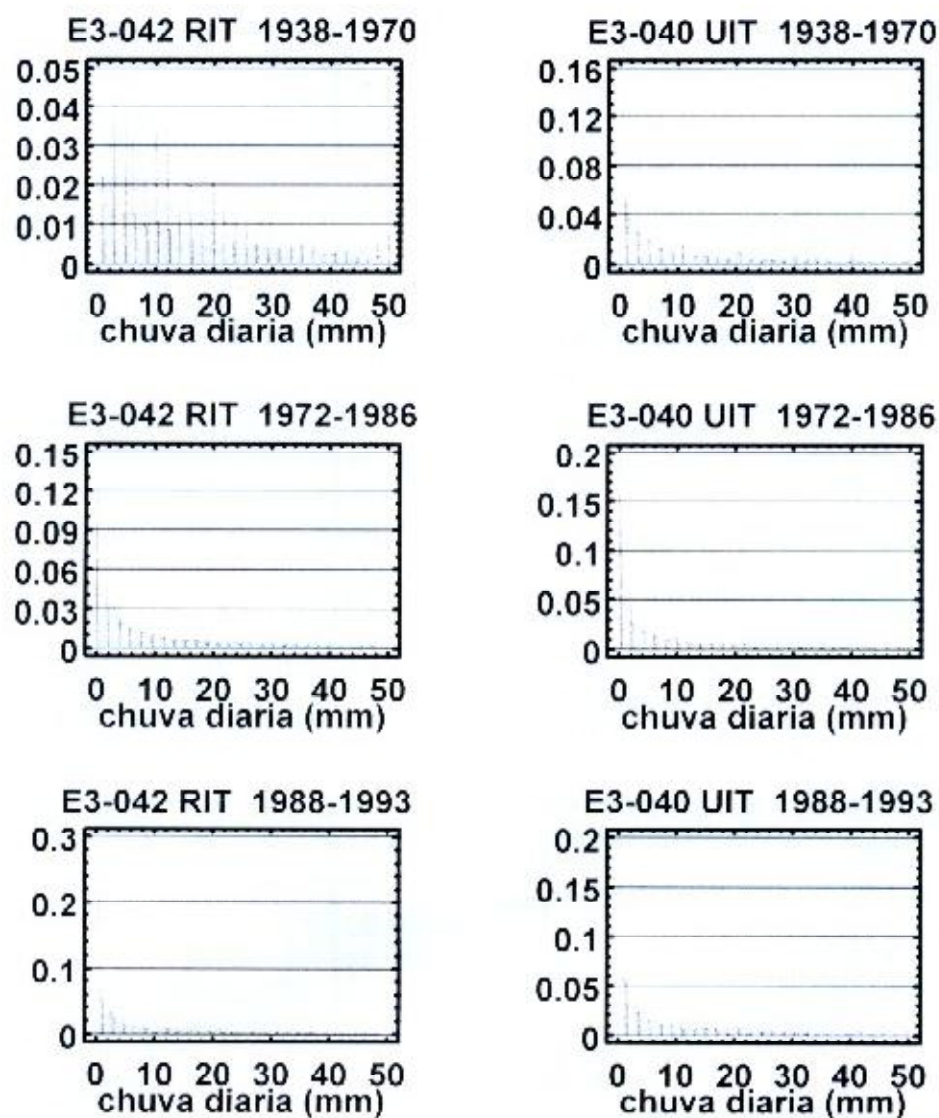


Figura 6. Histogramas da altura da chuva diária, para os Postos E3-042/Represa Itatinga e E3-040/Usina Itatinga (análogos períodos das figuras precedentes).

negligência ou incompetência dos observadores, deve-se mais uma vez ressaltar o problema decorrente das transformações para milímetros de dados que eram expressados em polegadas (presumivelmente com aproximação até 0,1 pol.), no caso de alguns postos que originariamente eram operados pela "São Paulo Railway". Por outro lado, há a possibilidade de que alturas pequenas de chuva fossem desprezadas por motivos operacionais, ou seja, aquelas eventualmente julgadas de pouco interesse do ponto de vista dos engenheiros ferroviários responsáveis pela manutenção da linha férrea, conforme hipótese já anteriormente aventada. Num caso ou no outro, há uma perda de informação, especialmente no que concerne às chuvas diárias de valores mais baixos, dando origem a artefatos, tais como as "rupturas" observadas na Figura 1, entre 1970-72.

No período seguinte, 1972-1986, pelo contrário, com respeito à maior parte dos postos pluviométricos, os histogramas apresentam grande regularidade ou, então, são mínimas aquelas porventura observadas. Isso corresponde à fase de organização do serviço de hidrologia estadual do DAEE (Depto. de Águas e Energia Elétrica do Estado de São Paulo), iniciada em 1969-1970. Assim, com abundância de recursos materiais e humanos, tornou-se possível organizar e reaparelhar a rede, empreender um melhor treinamento dos observadores, bem como, exercer uma fiscalização e supervisão efetivas do seu trabalho de campo. Em consequência, além do treinamento ministrado, havia condições de substituir imediatamente aqueles que não tivessem condições (ou disposição) para executarem as medições com um nível adequado de fidedignidade.

Para que se possa ter uma idéia quantitativa da melhoria conseguida, basta dizer que apenas 30% dos dados, medidos entre 1958 e 1968, foi publicado, desde que os dados restantes eram de tão baixa qualidade que tiveram de ser refugados. Este grau de aproveitamento aumentou para quase 80% com relação aos dados de 1969, tendo

atingido 90% para os anos de 1970 e seguintes.

Finalmente, no último período, a partir de 1988, há indícios, também extraídos da inspeção dos histogramas, de certa degradação na qualidade dos dados, para a maioria dos postos analisados; embora os defeitos não sejam tão graves quanto os verificados até 1970. Não se pode, é bem verdade, afastar a hipótese de que tais irregularidades, agora, sejam atribuíveis (pelo menos em parte) às flutuações de amostragem, que costumam tornarem-se mais evidentes no caso de amostras de menor tamanho. Contudo, sem dúvida, ocorreu no último período uma degradação da rede hidrométrica do DAEE, em termos operacionais, causada pela escassez de recursos. Note-se que um tipo de erro freqüente neste período refere-se a totais acumulados de chuva, em dias consecutivos, pela ausência de medições em sábados, domingos e feriados. De fato, a título de ilustração, diga-se que em boa parte dos postos localizados na Grande São Paulo não são feitas observações em fins de semana, devido à ausência (repouso semanal) ou à negligência dos observadores. O mesmo ocorre com respeito a feriados. Por outro lado, devido às carências mencionadas, têm ultimamente se tornado falhos, tanto o treinamento quanto a supervisão e a fiscalização do pessoal que opera os postos da rede estadual.

Cabe, aqui, uma observação final com respeito à metodologia utilizada, que de resto é extremamente simples. De fato, vimos como os gráficos de densidades empíricas (ou histogramas) conseguem revelar graves erros ou inconsistências existentes nos dados diários de chuva. Isso se aplicaria, igualmente, para dados horários ou ainda referentes a intervalos de tempo mais curtos (chuva de 1 min, de 10 min, etc.). Contudo, lembremos que um tal procedimento funciona, mais ou menos, como a conhecida "prova dos 9" usada como teste de consistência para operações aritméticas. Portanto, mesmo que os histogramas não venham a revelar a

existência de distorções com respeito ao padrão básico, ao longo dos vários períodos examinados, isso não significa que não possam estar eventualmente embutidos nos dados alguns vícios. Assim, a informação extraída dos histogramas necessita ainda ser cruzada com outros informes, envolvendo o histórico de cada posto e a análise dos procedimentos de controle de qualidade correntemente utilizados.

Por outro lado, note-se que erros eventualmente ocorridos nas medidas referentes às chuvas de maior intensidade não conseguiriam ser reveladas neste contexto. Além disso, como se tratam de valores extremos superiores, os erros não alterariam perceptivelmente as formas geométricas dos histogramas, uma vez que são os valores responsáveis apenas pelo delineamento das suas caudas. De fato, mesmo que os erros fossem substanciais, dificilmente seriam observáveis, ainda que as caudas não tivessem sido truncadas nos gráficos. Para o controle de qualidade de dados extremos, sem dúvida, seria necessário lançar mão de técnicas específicas. No caso de valores extremos resultantes de sistemas de chuvas generalizadas, de grande intensidade, erros de medida dessas chuvas extremas poderiam eventualmente serem detectados via a comparação das alturas das chuvas em pontos vizinhos; embora se deva ter em mente a grande variabilidade espacial da precipitação e, além disso, a influência dos efeitos topográficos, os quais podem explicar grandes diferenças nas alturas pluviométricas em locais espacialmente próximos.

Um ponto a discutir refere-se ao problema da qualidade dos dados mensais. É uma pergunta que logo pode ser colocada é a seguinte: os erros e inconsistências nos dados diários refletem-se, necessariamente, nos dados mensais da chuva? Sem dúvida, erros fortuitos nos dados diários, que possam ser descritos probabilisticamente através de uma lei simétrica e de média zero, com variância não muito grande, podem se compensar estatisticamente e não afetarem de maneira significativa a qualidade das

medidas mensais de chuva. Também erros resultantes dos acúmulos de chuva nos dias de fins de semana, igualmente, não costumam afetar os dados mensais, a menos daqueles fins de semana que envolvam uma virada de mês. Contudo, em princípio, no caso de um longo período em que se multipliquem erros grosseiros ou graves inconsistências nos dados diários, sem dúvida, os dados mensais respectivos passam a ser considerados altamente suspeitos. Quanto a mudanças reais no comportamento da chuva mensal, de longo período, talvez possam se refletir nos parâmetros das leis de probabilidade para a chuva diária; porém, mais dificilmente, no sentido de introduzirem irregularidades flagrantes nos histogramas. Pelo contrário, alterações na chuva podem ser detectadas de uma maneira mais incisiva, através do comportamento da chuva diária, antes que possam se refletir sob a forma de alterações da precipitação mensal; principalmente, no caso das que se restrinjam às chuvas diárias de pequena intensidade.

Deve-se mencionar que os problemas relativos à qualidade da pluviometria diária, por nós detectados, não se referem apenas aos postos na Baixada Santista e no alto da Serra do Mar (numa área de possível influência do parque industrial de Cubatão), aqui analisados, pois nossa experiência aponta para também ocorrerem com respeito a um grande número de outros postos pluviométricos da rede estadual paulista. Porém, há um caso merecendo menção especial, que é o posto E3-036 (Estação da Luz-São Paulo), que também foi operado pelos ingleses. Trata-se de uma longa série pluviométrica diária, que se inicia em janeiro de 1888; portanto, uma série secular cujos dados, se fossem absolutamente confiáveis, ganhariam uma importância ímpar para efeito de estudos climáticos. Ora, podemos apenas adiantar que de um exame preliminar expedito das planilhas para os dados diários da "Estação da Luz" no período 1888-1935, por simples inspeção, fica revelado que o fenômeno ligado ao problema de conversão de unidades existe, pelo menos entre os anos

1888 e 1907, embora não aparentemente nos anos seguintes. Sem dúvida, os dados desta série pluviométrica e de outras, para o Estado de São Paulo, mereceriam uma análise detalhada e aprofundada, "a posteriori".

Em particular, tudo isso implica em estritas providências que devem ser tomadas, quanto à análise da qualidade e a revisão dos dados diários de chuva, no contexto de toda pesquisa que exija sejam esses dados utilizados, em especial quando se tratar de estudos relacionados com mudanças climáticas, para os quais torna-se mais crítico o problema de erros ou de inconsistência nos dados. Assim, é possível que estudos realizados sem tais cuidados prévios fiquem irremediavelmente comprometidos pelos erros intrínsecos aos dados disponíveis.

Por outro lado, a necessidade de tais cuidados não se refere apenas ao caso dos dados pluviométricos da rede paulista, pois em princípio este mesmo tipo de problema deve ser corrente com respeito à maioria das redes meteorológicas ou pluviométricas dos vários estados brasileiros, ou mesmo fora do país; para as quais se recomenda sejam empreendidas, sistemática e preventivamente, tais análises de qualidade dos dados respectivos. Isso deve ser feito para um extenso período de tempo, ao longo de sub-períodos consecutivos, tendo em vista que a qualidade dos dados pode ser variável nesses sub-períodos, como ocorreu no nosso caso. Por outro lado, podendo também ser empreendido com respeito aos dados diários considerados ao longo de todo o ano ou, então, restritos a cada semestre ou trimestre do ciclo hidrológico. Tal proceder pode ser tornar apropriado quando houver a possibilidade de uma variação sazonal na idoneidade das medidas efetuadas. No caso do Nordeste brasileiro corre aliás a notícia, de caráter "folclórico", sobre observadores que descuravam das medidas de chuva durante o semestre seco porque, no seu entender, não ofereceriam interesse.

Sobre os procedimentos inerentes ao contexto da Figura 1, isto é, o emprego de gráficos que se refiram às chuvas diárias de

menor intensidade, esses podem também assumir grande utilidade e se tornarem muito instrutivos, como aliás o foram no âmbito desta pesquisa. Ademais, gráficos dessa natureza, porém relativos a chuvas mais intensas, provavelmente terão sua utilidade. Além de outras técnicas referidas na "Metodologia".

Finalmente, cabe lembrar que tais providências para uma avaliação da qualidade dos dados da pluviometria, inscreve-se em um dos aspectos contidos na recomendação sobre a necessidade de medidas a serem tomadas para uma "gestão integrada da quantidade e da qualidade dos recursos hídricos, considerando as fases meteórica, superficial e subterrânea do ciclo hidrológico", conforme expressado na "Carta do Rio de Janeiro", ABRH (1991).

Mais uma vez se insiste, em princípio, para os indícios de degradação na qualidade dos dados a partir de 1986-1988, apesar de que parte das irregularidades nos histogramas possam ser decorrentes do processo de amostragem (naturais de ocorrerem quando de amostras de menor tamanho). Note-se que, se as irregularidades atribuíveis às deficiências para a manutenção da rede, sem falar na necessidade de sua expansão e modernização, virem ainda mais a se agravarem, sem dúvida isso conduziria a uma perda irreparável da qualidade de nossa "memória hidrometeorológica". Nesse caso, sem que sejam tomadas as iniciativas indispensáveis e urgentes na esfera governamental para corrigir tal situação, graves consequências poderiam advir no sentido do comprometimento da pesquisa científica, pela não disponibilidade de dados confiáveis e, também, com respeito a uma falta de legitimação dos trabalhos técnicos rotineiros, nas áreas de engenharia hidráulica, meteorologia, agricultura, etc.

Consta, a esse respeito, que algumas medidas estão de fato sendo implementadas, especialmente no sentido da automatização das redes hidrométricas brasileiras, de sorte que se possa dispor de dados confiáveis, imunes a erros humanos de medição, além de

sua imediata disponibilidade "on-line". Com efeito, o funcionamento de uma rede em moldes tradicionais exige, nos dias de hoje, a utilização de recursos humanos e financeiros que, provavelmente, a sociedade não mais suportaria. Decerto, a implantação de redes telemétricas pede, de início, a aplicação de investimentos de algum peso em termos financeiros, embora facilmente justificáveis a médio e a longo prazo em termos do seu custo-benefício.

AGRADECIMENTOS

Agradecemos ao CNPq e à FAPESP pelo apoio a esta pesquisa, em várias de suas etapas. Bem como, ainda, ao DAEE-CTH, pela presteza na cessão dos dados de chuva utilizados em nossas pesquisas.

REFERÊNCIAS

- ABRH-Associação Brasileira de Recursos Hídricos (1991), "Carta do Rio de Janeiro", aprovada no VIII Simpósio Brasileiro de Recursos Hídricos; ver em : *Catálogo ABRH (1977-1995)*, ABRH, São Paulo, 1995, pp. 31-34.
- ANANTHAKRISHNAN, R. & SOMAN, M.K. (1989), Statistical Distributions of Daily Rainfall and its Association with the Coefficient of Variation of Rainfall Series, *Intern. J. Climatology*, Vol. 9, pp. 485-500.
- EASTON, G.S. & McCULLOCH, R.E. (1990), A Multivariate Generalization of Quantile-Quantile Plots, *Journal of the American Statistical Association*, Vol. 85, pp. 376-386.
- FREEMAN, J.M. (1984), Two-Phase Regression and Goodness of Fit, *Commun. Statist.-Theor. Meth.*, Vol. 11, pp. 1321-1334.
- HINKLEY, D.V. (1969), Inference About Intersecion in Two-Phase Regression, *Biometrika*, Vol. 56, pp. 495-504.
- HINKLEY, D.V. (1971), Inference in Two-Phase Regression, *Jour. American Statist. Assoc.*, Vol. 55, pp. 736-743.
- JURAS, J. (1994), Some Common Features of Probability Distributions for Precipitation, *Theor. Appl. Climatol.*, Vol. 49, pp. 69-76.
- KUTIEL, H. (1990), Variability of Factors and Their Possible Application to Climatic Studies, *Theor. Appl. Climatol.*, Vol. 42, pp. 169-175.
- MIELKE Jr., P. W. (1973), Another Family of Distributions for Describing and Analyzing Precipitation Data, *Journ. Applied Meteorol.*, Vol. 12, pp. 275-280.
- LAVERY, B., KARIKO, A. & NICHOLLS, N. (1992), A Historical Rainfall Set for Australia, *Australian Met. Mag.*, Vol. 40, pp. 33-39.
- SOLOW, A.R. (1987), Testing for Climate Change: An Application of the Two-Phase Regression Model, *Journ. Climate Appl. Meteorol.*, Vol. 26, pp. 1401-1405.
- SOMAN, M.K. & KUMAR, K.K. (1990), Some Aspects of Daily Rainfall Distributions Over India During the South-West Monsoon Season, *Intern. J. Climatology*, Vol. 10, pp. 299-311.
- SUBBARAMAYYA, I. & RAO, N.J. (1964), The Frequency Distribution of Rainfall of Different Intensities, *Journ. Met. Soc. Japan*, Vol. 42, pp. 277-284.
- SUZUKI, E. (1964-65), Hyper Gamma Distribution and Its Fitting to Rainfall Data, *Papers in Meteorol. Geophys.*, Vol. 15, pp. 30-51.
- XAVIER, T. de Ma. B. S., XAVIER, A. F. S. & SILVA DIAS, M. A. F. da (1994a), Evolução da Precipitação Diária num Ambiente Urbano: O Caso da Cidade de São Paulo, *Rev. Bras. Meteorologia*, Vol. 9, pp. 44-53.
- XAVIER, T. de Ma. B. S., SILVA DIAS, M.A.F. da & XAVIER, A.F.S. (1994b), Evolução de Parâmetros da Precipitação Diária na Grande São Paulo (1933-1993), *Anais do VII Congr. Bras. Meteorologia*, Vol. 2, pp. 347-349.
- XAVIER, A.F.S., XAVIER, T. de Ma. B. S., SILVA DIAS, M.A.F. da & TOLEDO, J.R.C. de (1995), Análise da Qualidade de Séries Diárias da Pluviometria no Estado de São Paulo do Banco de Dados Hidrométricos do DAEE-CTH, em : Clarke, R.T., Chaudhry, F. & Baptista, M. (Editores), "Desenvolvimento Sustentável dos Recursos Hídricos", Volume 1, pp. 325-330, Artigos Selecionados do XI Simpós. Bras. de Recursos Hídricos/ II Simpós. de Hidráulica dos Países de Língua Oficial Portuguesa, Recife-PE, 5-10 nov. 1996, ABRH-Associação Brasileira de Recursos Hídricos.
- XAVIER, T. de Ma. B. S., XAVIER, A.F.S. & SILVA DIAS, M.A.F. da (1996), Analyzing Changes of Daily Precipitation in the Great São Paulo and Need of Data Quality Checking, *Conference on Environmetrics in Brazil*, July 22-26, 1996, São Paulo, *Abstracts* (Institute of Mathematics and Statistics-University of São Paulo), pp. C4-C5.

- MILKS, D.S. (1990), Maximum Likelihood Estimation for the Gamma Distribution Using Data Containing Zeros, *Journal of Climate*, Vol. 3, pp. 1495-1501.
- MOOLHISER, D.A. & ROLDÁN, J. (1982), Stochastic Daily Precipitation Models, 2. A Comparison of Distributions of Amounts, *Water Resour. Res.*, Vol. 18, pp. 1461-1468.

Quality of Daily Rainfall Data in S.Paulo-Brazil

ABSTRACT

This paper concerns the analysis of the quality of daily rainfall records of the State of São Paulo-Brazil, from the rainfall data bank of the DAEE-CTH. We have detected several serious problems in the data until 1970; between 1972 and 1986, in general, the quality of the data can be considered good; from 1988 on, the analysis shows some deterioration, although the problem seems to be less serious than during the first period. As a consequence we suggest the analysis and the revision of all daily rainfall data used in any research, especially if they are taken in the context of studies on climatic changes.